Mobile genetic elements in *Methanobacterium thermoformicicum*

Promotor: Dr. W. M. de Vos
hoogleraar in de Bacteriële Genetica

Co-promotor: Dr. H. I. L. Eggen
universitair docent bij de vakgroep Microbiologie

J. Nölling

Mobile genetic elements in *Methanobacterium thermoformicicum*

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
Dr. H. C. van der Plas,
in het openbaar te verdedigen
op maandag 29 maart 1992
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen

Graag wil ik al diegenen die mij op welke wijze dan ook hebben geholpen bij het tot stand komen van dit proefschrift van harte bedanken.

This study was carried out at the Department of Microbiology, Bacterial Genetics Group, Wageningen Agricultural University, The Netherlands.

Part of the cover-illustration was taken from an advertisement of the computer company *boeder* published in: *Der Spiegel*, 34/1992, p. 118.

Stellingen

1. Gezien het grote aantal niet-gekarakteriseerde prokaryoten lijkt de indeling in *Archaea* en *Bacteria* door C. R. Woese en medewerkers van voorlopige aard.
   Woese et al. 1990. *Proc. Natl. Acad. Sci. USA* 287:4576-4579

2. Het niet in aanmerking nemen van DNA-modifikaties kan bij het in kaart brengen van prokaryotische genomen het trekken van verkeerde conclusies bevorderen.
   R. Stettler en T. Leisinger (1992). *J. Bacteriol.* 174:7227-7234.

3. De isolatie van thermofiele methanogenen uit riviersedimenten impliceert dat deze organismen wellicht beter als 'fakultatieve thermofielen' beschreven kunnen worden.
   Dit proefschrift, Hoofdstuk 2

4. Onverwachte resultaten zijn de meest interessante.

5. Vooroordelen tegenover vluchtelingen berusten op vluchtige indrukken.
   Amnesty International

6. Het aanwakkeren van een oorlogsstemming door nederlandse media tijdens bepaalde internationale voetbalwedstrijden is onaanvaardbaar.

7. Het valt moeilijk te verklaren dat experimentjes met plasmiedjes en faagjes soms leuker kunnen zijn dan rustigjes met een krantje en een pilsje in het zonnetje te zitten.

8. Verkorting van de studieduur bevordert niet de noodzakelijke 'cultuuroverdracht' (*de Volkskrant*, 09.01.93) aan de universiteit maar versterkt de trend tot opleiding van vakspecialisten met geïsoleerde kennis.

9. De aanpak van milieuproblemen wordt belemmerd omdat met onaangename waarheden geen verkiezingen te winnen zijn.

10. Het is aan te bevelen om met het formuleren van stellingen op tijd te beginnen.

Stellingen behorende bij het proefschrift "Mobile genetic elements in *Methanobacterium thermoformicicum*" van Jörk Nölling.

Wageningen, 29 maart 1993.

Stellingen

1. Gezien het grote aantal niet-gekarakteriseerde prokaryoten lijkt de indeling in *Archaea* en *Bacteria* door C. R. Woese en medewerkers van voorlopige aard.
   Woese et al. 1990. *Proc. Natl. Acad. Sci. USA* 287:4576-4579

2. Het niet in aanmerking nemen van DNA-modifikaties kan bij het in kaart brengen van prokaryotische genomen het trekken van verkeerde conclusies bevorderen.
   R. Stettler en T. Leisinger (1992). *J. Bacteriol.* 174:7227-7234.

3. De isolatie van thermofiele methanogenen uit riviersedimenten impliceert dat deze organismen wellicht beter als 'fakultatieve thermofielen' beschreven kunnen worden.
   Dit proefschrift, Hoofdstuk 2

4. Onverwachte resultaten zijn de meest interessante.

5. Vooroordelen tegenover vluchtelingen berusten op vluchtige indrukken.
   Amnesty International

6. Het aanwakkeren van een oorlogsstemming door nederlandse media tijdens bepaalde internationale voetbalwedstrijden is onaanvaardbaar.

7. Het valt moeilijk te verklaren dat experimentjes met plasmiedjes en faagjes soms leuker kunnen zijn dan rustigjes met een krantje en een pilsje in het zonnetje te zitten.

8. Verkorting van de studieduur bevordert niet de noodzakelijke 'cultuuroverdracht' (*de Volkskrant*, 09.01.93) aan de universiteit maar versterkt de trend tot opleiding van vakspecialisten met geïsoleerde kennis.

9. De aanpak van milieuproblemen wordt belemmerd omdat met onaangename waarheden geen verkiezingen te winnen zijn.

10. Het is aan te bevelen om met het formuleren van stellingen op tijd te beginnen.

Stellingen behorende bij het proefschrift "Mobile genetic elements in *Methanobacterium thermoformicicum*" van Jörk Nölling.

Wageningen, 29 maart 1993.

Contents

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Nölling, Jörk

Mobile genetic elements in *Methanobacterium thermoformicicum* / Jörk Nölling. - [S.1. : s.n.]. - Ill.
Thesis Wageningen. - With ref. - With summary in Dutch.
ISBN 90-5485-075-2
Subject headings: molecular genetics / *Methanobacterium thermoformicicum.*

General introduction

The diversity of life as the result of evolution from a common origin has generated considerable interest in the classification of organisms. The accuracy, however, with which a phylogenetic tree reflects the relationships among the organisms strongly depends on the markers and methods used for comparison. In general, markers that are directly related to the nucleotide sequences of a genome are more informative for the determination of phylogenetic relationships than are phenotypic properties (157). A nevertheless meaningful phenotypic classification was the division of the living world into the major groups of prokaryotes and eukaryotes which initially was based on readily detectable phenotypic properties such as the pronounced differences in their cellular organization. Only later this distinction was confirmed by the identification of comparable molecular characteristics such as differences in ribosome structure (159). However, the prokaryote-eukaryote dichotomy was based on the comparison of eukaryotic markers with those derived from the model bacterium *Escherichia coli* that was considered as a typical representative of the prokaryotic kingdom (159).

Compared to the phenotypic diversity of eukaryotes, the prokaryotes constitute a group of phenotypically relatively homogeneous organisms. This led to the assumption that the prokaryotes have a monophyletic origin, i.e. are descendants of a single common ancestor. Despite the lack of complex phenotypic markers, prokaryotic systematics was mainly based on comparative morphology of cells or colonies, differences in cell-walls revealed by Gram-staining and readily detectable physiological properties such as growth with or without oxygen, nitrogen fixation or photosynthesis (157). However, those criteria have only limited value for prokaryotic taxonomy mainly for two reasons (157). Firstly, lateral gene transfer or adaptation to environmental stress may have resulted in similar phenotypes of two distantly related organisms. And secondly, the evaluation of a phenotypic marker as phylogenetically significant is often a subjective one (157).

More reliable information on the prokaryotic phylogeny was obtained from analysis of various molecular markers. Cell wall analysis and immunological studies, together with initial approaches in genotype-based classification such as comparison of the DNA base ratios (GC-content) and analysis of the total DNA relatedness, were applied in the phylogenetic characterization of lower taxa (157). A considerable step forward was the use of macromolecules as 'evolutionary clocks' to determine evolutionary distances (72). Macromolecules such as nucleic acids and proteins consist of a large number of independent constituents, i.e. nucleotides and amino acids, whose sequential order provides the information for their function. Since the number of possible different sequences is extremely high, two sequences with extensive similarity are most likely evolutionary related rather than the product of convergent evolution. As the result of continuously occurring, selectively neutral mutations, the observed accumulated differences between two sequences are analogous to a relative time span, i.e. the sequence differences can be used to calculate the evolutionary distance of the compared sequences from a common ancestor. Among the sequences that may be used as molecular chronometers for phylogenetic analysis of higher taxa, most suitable ones are those that have been established in early stages of cell's

evolution. Among those, ribosomal RNAs are extremely suitable since (i) they are functionally highly constant and universally distributed, (ii) they contain enough information to be statistically reliable, (iii) they contain both moderately and highly conserved regions and, therefore, are suitable for measuring both small and large phylogenetic distances, and finally (iv) their isolation and analysis is relatively easy (159).

Based on comparative analysis of 16S rRNA sequences, Woese and coworkers discovered a basic division of the known prokaryotic organisms that led them to propose a third primary lineage distinct from eubacteria and eukaryotes, designated archaebacteria (Figure 1). A similar classification was established by comparison of other molecular chronometers with the capacity for high-taxa-resolution such as for instance components of the DNA-dependent RNA polymerase (74, 174), translation factors (20, 47) and ATPase components (37, 60, 62). Certain group-specific features could be identified on the molecular genetic and biochemical level (for review: 173) which confirmed the distinction between the three 'ur-kingdoms' of life. Since the new, genotype-based taxonomy should be reflected in terminology, Woese and coworkers (1990) proposed to rename the three primary lineages into the 'domains' *Archaea*, *Bacteria* and *Eucarya* (Figure 1).

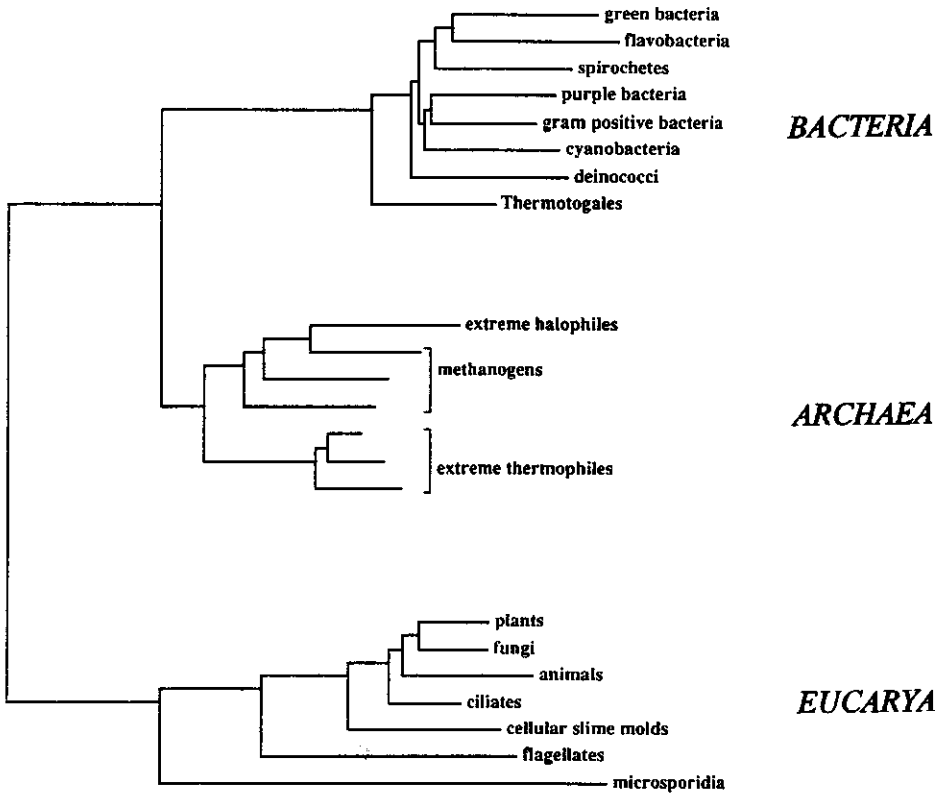


Figure 1. Unrooted universal phylogenetic tree (taken from Woese [159]), based upon 16S rRNA sequence comparison, showing the known three domains of life.

Besides certain unique archaeal characteristics as ether-linked lipids (26), unusual modified tRNA-nucleotides (30) and the methanogenic C_1-metabolism (27, 68), most of the features that allow to distinguish the *Archaea* from the *Bacteria* and the *Eucarya* are associated with the genetic machinery (173). The archaeal genomes appear to be composed of a single circular chromosome similar to the genome of their bacterial counterparts. As in *Bacteria*, archaeal genes are - with the exception of certain genes from halophilic organisms (10) - preceded by Shine-Dalgarno sequences and frequently are organized in transcriptional units (10). The majority of the archaeal features, however, were found to be more similar to those from the *Eucarya* such as the presence of introns (73, 111) and 7S RNA (66), the structure of promoters (44, 48), and the sensitivity to certain antibiotics (165, 173). This striking similarity between *Archaea* and *Eucarya* is also reflected by sequence comparison of ribosomal proteins (156), translation factors (20, 47), components of the DNA-dependent RNA polymerase (74, 173), DNA ligases (75), components of the ATPase (37, 60, 62) and histone-like proteins (129). Although not reflected on 16S RNA-level, this archaeal-eucaryal similarity in basic molecular characteristics provides reasonable arguments for a closer relationship of the *Archaea* to the *Eucarya* than to the *Bacteria*. In contrast, the sequences of enzymes of the central metabolism as GAPDH (50), malate dehydrogenase (57), 3-phosphoglycerate kinase (33) and β-glucosides hydrolyzing enzymes (31) were found to be more related between *Bacteria* and *Eucarya* (173).

The domain *Archaea* comprises at least three major groups of prokaryotic organisms with, compared to their bacterial counterparts, unusual phenotypes: the extreme thermophilic, sulfur-dependent species, the extreme halophiles and the methanogens (159). Phylogenetically, the *Archaea* can be subdivided into two main lineages (kingdoms): the *Crenarchaeota* (158) which comprise most of the hyperthermophiles, and the *Euryarchaeota* (158), a phenotypically more diverse collection of organisms including the three groups of methanogens, the extreme halophiles, the extreme thermophilic genera *Thermococcus* and *Thermoplasma* and the sulfate-reducing *Archaeoglobus*. The branching order of the euryarchaeotal tree suggests that this kingdom has evolved from an ancient sulfur-reducing, hyperthermophilic organism via the methanogenic to the extreme halophilic organisms (159). Moreover, it is assumed that an extreme thermophilic, sulfate-reducing organism has been the common ancestor of the *Archaea* since this phenotype is the only one shared by both archaeal kingdoms (158).

Methanogens are, with respect to morphology and growth requirements, a diverse group which obtain their energy for growth from the conversion of low molecular weight compounds to methane (for review: 154). The main products of fermentative decomposition of organic matter, H_2 and CO_2, formate, and acetate are the major substrates of methanogenic bacteria. In addition, other C_1-compounds and certain alcohols may serve as substrates. Methanogens exist as free-living and endosymbiotic organisms (150) and occupy a wide range of anaerobic habitats in soil and aquatic environments, the intestines of certain organisms, bioreactors, and extreme environments such as hydrothermal vents and hypersaline lakes.

The biotechnological importance of methanogenic bacteria is reflected by their application in biodegradation of waste and the simultaneous production of biogas as an alternative form of energy and their potential to convert xenobiotics (2, 19, 53, 77). In addition to these (potential) applications, the scientific interest in molecular biology of methanogenic bacteria as representatives of the *Archaea* accounts for the progress in the genetics of methanogens which was subject of various recent reviews (10, 122, 171).

MOBILE GENETIC ELEMENTS

Prokaryotic chromosomes are regarded to consist of a set of essential genes which are interspersed with a variety of sequences including accessory genes, noncoding segments and (short) repeated sequences (80). Retaining a basic stability of the genomic organization, the chromosomal DNA is subjected to several genetic changes including point mutations and small and large scale rearrangements such as insertions, deletions and inversions. In addition to those changes of the resident genetic material, horizontal gene transfer serves as an important source for genomic variations which allow an adaptation of the host to environmental stress conditions. As a consequence of the uptake and assimilation of foreign DNA, the chromosome can be viewed as a mosaic, comprising recipient-derived DNA interspersed with DNA from other sources (141).

A comparable organization and dynamic nature as found for the chromosomal DNA is in principle also found in the smaller-sized genomes of plasmids and phages. Therefore, elucidation of the genomic organization of these elements may provide insight into the structure of genomes in general. Moreover, plasmids and phages belong, together with insertion sequences (IS), to a class of accessory elements that force the change of the cell's genomic constitution: their movement within or between prokaryotic genomes often is accompanied with genomic rearrangements (promoted by insertion elements) and the horizontal transmission of genetic material (plasmid- and phage-mediated). Thus plasmids, phages and insertion elements which together constitute a family of mobile genetic elements (79), appear to play a significant role in evolutionary processes.

The naturally occurring IS elements comprise defined DNA sequences ranging from 0.8 to 3 kb and are joined to autonomously replicating molecules (replicons) such as the prokaryotic chromosome, phage- and plasmid DNA. A characteristic of these mobile elements is that they specify functions that allow them to move from one site to another, either intra- or intermolecularly. Different types of genomic changes are promoted by IS elements (35): (i) transposition of IS elements itself may cause disruption of coding regions or may lead to a fusion of replicon, (ii) the presence of several genomic copies of IS elements provide sites for reciprocal recombination which results in genomic rearrangements such as deletions and inversions, and (iii) two IS elements flanking genomic-derived genes may form a composite transposon thereby mobilizing the internal genomic sequences.

Plasmids and phage genomes are, in contrast to IS elements, autonomously replicating genetic elements that either exist extrachromosomally or - as the result of site-specific integration or homologous recombination - may become part of the host chromosome. While the mobility of IS elements is restricted to the genome of a single cell, many plasmids and phages have the potential to mediate the transfer of their genomes from one cell to another. This transmission of DNA can be accompanied by mobilization of host DNA, including IS elements, by conjugation (plasmid-directed) and transduction (phage-directed). In addition to natural transformation, i.e. the uptake and incorporation of exogenous DNA, conjugation is regarded as the major mechanism for horizontal DNA transfer in nature that occurs between phylogenetically divergent prokaryotes (1, 87) and may even bridge the boundaries between members of different domains, including DNA transfer between pro- and eukaryotes ('interkingdom sex') (1, 49, 141).

The study of genetic mobile elements may therefore contribute to the understanding of genomic organization and processes that influence the genetic constitution of prokaryotic organisms. Moreover, in particular plasmids are useful model systems for the investigation of the fundamental process of DNA replication and may serve as basic tools in recombinant

DNA technology.

ARCHAEAL MOBILE ELEMENTS

Plasmids

Plasmids have been identified in all three major groups of *Archaea* (for review: 10; Table 1).

Among the hyperthermophilic, sulfur-dependent *Archaea* only two plasmids, designated pGT5 (32) and pBG from *Sulfolobus* strain B6 (143), have been described whereas two other extrachromosomal elements, pSB12 and pSL10, represent lysogenic forms of viruses or virus-like particles that will be discussed in the next section. The small, 3.45-kb multicopy plasmid pGT5 is harbored by a recently isolated hyperthermophilic archaeon (strain GE5) which has an optimal growth temperature of 94°C (32). Surprisingly, pGT5 was found to be nearly relaxed at physiological temperatures (95°C) (11), although strain GE5 contains reverse gyrase activity that is common to extreme thermophilic organisms of both archaeal and bacterial origin (8, 9). This activity was believed to be necessary for thermostabilization of closed circular DNA by introduction of positive supercoils. It is supposed that the relaxed state of pGT5 at physiological temperatures *in vivo* is the result of both reverse gyrase activity and the presence of nucleosome-like structures and might be representative for the topology of chromosomal and extrachromosomal DNA of hyperthermophiles (11, 34).

Most progress in biology of archaeal plasmids has been made with those from the extreme halophilic organisms. For a part this is due to the large number of plasmids that have been described for these organisms, ranging in size from small, 1.7-kb high copy number plasmids up to 690-kb megaplasmids (Table 1), and the established transformation systems (see below). Many halophilic bacteria contain heterogeneous plasmid populations and a considerable part of the genomic information is located on extrachromosomal elements (14, 40, 112, 127, 145). The 150-kb pHH1 has been identified as the main plasmid of *Halobacterium halobium* (153) (several strains of *H.halobium* and *H.cutirubrum* probably belong to the species *H.salinarium* according to Tindall [1992]). Plasmid pHH1 contains a large number of different halobacterial IS elements that cause insertion, deletions and rearrangements of pHH1 (115, 118, 119). Other *Halobacterium* species were found to contain structural derivatives of pHH1, either as covalently closed DNAs or integrated into the chromosome (28, 112). Initial observations of an correlation between the ability to form gas vesicle and the presence of plasmids (137, 153) has led to the identification of *vac* genes located on the plasmids pHH1 (58) and pNRC100 (23, 64). Detailed analysis of the 200-kb pNRC100 resulted in the identification of two large (35 to 38 kb) inverted sequences which are supposed to mediate inversion of the intervening single-copy regions, one of which contains the *vac*-region, and constitute a possible regulation mechanism for gene expression (100). In addition, the origin of replication has been mapped for plasmid pNRC100 (100), pHH1 (119) and pHK2 (56). A function could also assigned to plasmid pΦHL, a deletion derivative of the ΦH prophage (133) whose nucleotide sequence and transcriptionally active regions have recently been determined (38). *H.halobium* R_1 cells carrying pΦHL become resistant to infection by phage ΦH due to a immunity-conferring coliphage-like repressor encoded by pΦHL (69).

Treatment of *H.halobium* GRB with the gyrase inhibitor novobiocin resulted in the

Table 1. Archaeal plasmids

Plasmid	Host	Size (kb)	References for Relevant Characteristics		
			Isolation/Sequence[a]	Function[b]	Transformation[c]
pBG[d]	*Sulfolobus* strain B6	35	(143)	–	–
pGT5	hyperthermophilic, sulfur-metabolizing isolate GE5	3.5	(32)	–	–
pNRC100	*Halobacterium halobium* NRC-1[e]	200	(23)	*vac* genes (23, 64)	–
pHH1	*H.halobium* NRC817	143	(153)	*vac* genes (58)	(6, 18)
pHH2	*H.halobium* DSM670	152	(112)	–	–
pHH3	*H.halobium* DSM671	76	(112)	–	–
pHSB1	*H.halobium* SB3	1.7	(28/41)	Rep protein (61)	(41, 85)
pHGN1	*H.halobium* GN101	1.8	(28/45)	Rep protein (61)	–
pGRB1	*H.halobium* GRB	1.8	(28/42)	Rep protein (61)	(17)
pΦHL	*H.halobium* R_1L	19	(133/38)	ΦH immunity (133, 69)	–
pHT1	*H.trapanicum*	122	(112)	–	–
pHT2	*H.trapanicum*	6	(112)	–	–
pHC1	*H.cutirubrum*	122	(112)	–	–
pB	*Halobacterium salinarium* 5	41	(136)	–	–
pRDS102	*H.salinarium* 5	69	(136)	*vac* genes (136, 137)	–
pD	*H.salinarium* 5	134	(136)	–	–
pHV1	*Haloferax volcanii* DS2	86	(112)	–	–
pHV2	*H.volcanii* DS2	6.4	(112/12)	–	(12, 84, 101)
pHV3	*H.volcanii* DS2	440	(14)	–	–
pHV4	*H.volcanii* DS2	690	(14)	–	–
pHV11	*H.volcanii* WR11	13	(126)	–	–
pHV12	*H.volcanii* WR12	25	(126)	–	–
pHV13	*H.volcanii* WR13	44	(126)	–	–
pHK2	*Haloferax* Aa2.2	10.5	(54)	–	(54, 56, 84, 101)
pHM1	*Halococcus morrhuae* CCM 537	48	(96)	–	–
pHM2	*H.morrhuae* CCM 537	6	(96)	–	–
pC2A	*Methanosarcina acetivorans* C2A	5.1	(140)	–	–
pURB500	*Methanococcus* sp. nov.	8.7	(160)	–	–
pURB800	*Methanococcus jannaschii*	64	(166)	–	–
pURB801	*M.jannaschii*	18	(166)	–	–
pURB900	*Methanococcus* AG86	20	(166)	–	–
pMP1[d]	*Methanolobus vulcani*	6.9	(144, 155)	–	–
pT3	*Methanobacterium* sp. nov.	7.3	(5)	–	–
pME2001	*M.thermoautotrophicum* Marburg	4.4	(88/7)	–	–
pFV1	*M.thermoformicicum* THF	13.5	(103/106)	*Mth*TI system (104)	–
pFZ1	*M.thermoformicicum* Z-245	11	(103/106)	*Mth*ZI system (105)	–
pFZ2	*M.thermoformicicum* FTF	11	(103)	*Mth*FI system (105)	–

[a] Complete nucleotide sequence.

[b] Identified functions (based on experimental data or on homology with known proteins), including plasmid replication- and maintenance functions.

[c] Transformation of the natural host or close relatives using the original plasmid or engineered derivatives including shuttle vectors.

[d] Possibly represents a prophage.

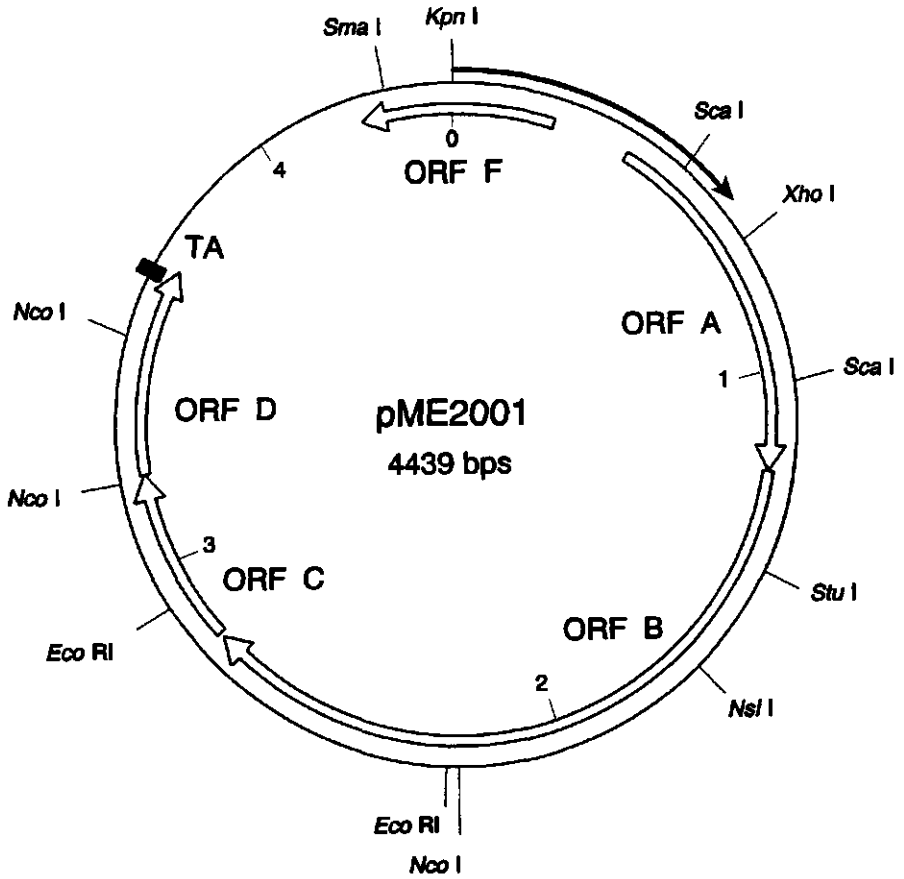
[e] Several strains of *H.halobium* and *H.cutirubrum* belong to the species *H.salinarium* (146).

accumulation of single stranded forms of the small 1.8-kb multicopy number plasmid pGRB1 suggesting that this plasmid replicates via a rolling circle mechanism (39, 139). Since sequence analysis of pGRB1 and the similar sized plasmids pHSB1 and pHGN1 revealed that each plasmids could code for a homologous protein of approximately 35 kDa (41, 42, 45), these putative proteins may be involved in a common mode of plasmid replication (42). This hypothesis was recently supported by Ilyina and Koonin (61), who found that the plasmid-encoded proteins of pGRB1, pHSB1 and pHGN1 share a conserved sequence motif with rolling circle replication initiation proteins.

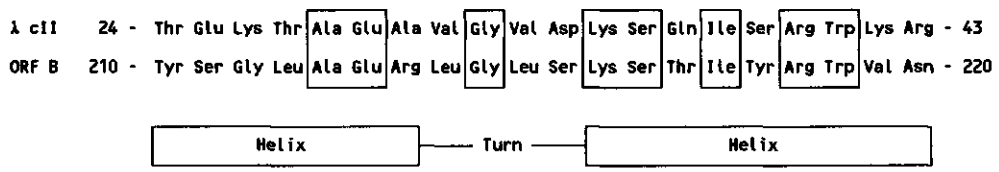
The application of genetic engineering techniques in halophilic members of the archaeal domain became possible with the development of efficient transformation systems. A natural mating system (94, 125) and a protoplast fusion technique (126) have been described for *Haloferax volcanii*. The polyethylene glycol mediated spheroplast transformation method, however, is the most commonly applied one and transformation protocols have been developed for the phylogenetically distant *H.halobium* (15, 41), *H.volcanii* (12, 17) and members of the genus *Haloarcula* (18). The vector plasmids used for transformation studies were derivatives of small halobacterial replicons equipped with a marker gene and, more recently, were expanded to halobacterial-*E.coli* shuttles vectors (6, 56, 84, 85). Selective markers used included genes conferring resistance to novobiocin (54) or mevinolin (84), or encoding a phenotypically selectable marker, such as gas vesicle synthesis (6). Expression of the bacterio-opsin (*bop*) gene in a *bop* mutant of *H.halobium* was observed after transformation with a pGRB1-based multicopy vector that carried an intact *bop* gene (82). Using a mutated halobacterial gyrase B gene which confers resistance to novobiocin (55) as selective marker, a cloning vector was constructed based on the pHK2 replicon of *Haloferax* Aa2.2 (54). Subcloning into an *E.coli* vector yielded the shuttle-vector pMDS1 which, however, could only stably maintained in *Haloferax* if propagated in a *dam E.coli* strain since methylated adenine residues are subject to restriction in the halophilic host (56). A derivative of pMDS1 has recently been used for expression of a alkaline serine protease, termed halolysin, from an unidentified halophilic strain in *H.volcanii* (67). Two other halobacterium-*E.coli* shuttle vectors, pUBP2 (derived from pHH1 of *H.halobium*; 6) and pWL102 (based on the *H.volcanii* pHV2 replicon; 84) were found to stably transform members of different halophilic genera (18) and may be useful as generalized cloning vectors in halophilic *Archaea*. The pHV2-derived vector has been used in complementation experiments to analyze the minimal DNA region necessary for gas vesicle synthesis (59). Recently, transformation of *H.halobium* was applied for specific *in vivo* rRNA mutagenesis of its single-copy rRNA operon (85). This technique makes use of a shuttle vector pHRZH, a derivative of pHSB1 from *H.halobium* SB3 (28) which is equipped with a mutated, *H.halobium* rRNA operon that confers resistance to two different antibiotics. After transformation, the sensitive, single-copy wild-type rRNA operon of the recipient is replaced by the vector-borne one via recombination, resulting in a transformed strain that directs synthesis of exclusively mutated ribosomes (85).

Representatives of all three major lines of the methanogenic branch of the *Archaea* were found to harbor plasmids (Table 1) ranging in size from 4.4 kb for the high-copy number plasmid pME2001 from *Methanobacterium thermoautotrophicum* Marburg (88) up to 64 kb for pURB800 from *Methanococcus jannaschii* (166). However, compared to what is known about plasmids from halophilic bacteria, only little progress has been made in plasmid biology of methanogens. This may be attributed to the fact that, compared to the halophilic *Archaea*, (i) plasmids are less common among methanogens, (ii) all plasmids identified so

A



B



C

λ cII (P_{RE}) GCCTCGTTGCGTTTGTTTGCACGAAC

TA1 3616 - TTTATTTTGATTTTTTTAGATTTTTTTAGATTTTTTTCTATC - 3575

TA2 3685 - ATTTTCTTGTATTTTTTTGCTTTCTCA - 3659

TA3 3771 - TTTTATTTGCATTTTTTTGCATTTTCTTGTATTTTTTTGCTTTCTCA - 3725

Figure 2. (A) Physical map and coding regions of plasmid pME2001 of *M.thermoautotrophicum* Marburg. ORFs A through F are indicated by open arrows whereas the mapped 611-bp transcript (91) is shown as a black arrow. The putative ORF B protein binding region (TA) is denoted by a solid block. Sizes are given in kilobases. (B) Comparison of the helix-turn-helix motif of the λ cII activator protein (51) with that deduced from ORF B. Identical amino acids are boxed. Numbers refer to the amino acid position. (C) Binding site of the λ cII protein (51) compared to the putative ORF B protein binding region of pME2001 (TA) composed of three sequence stretches (TA1 through TA3). Tetranucleotides similar to TTGC and spaced by six nucleotides are underlined. Numbers refer to sequence position according to Bokranz et al. (7).

far in methanogens are cryptic, i.e., no phenotype could be correlated with the presence of plasmids, and (iii) an efficient transformation system and suitable vectors are at present not available.

One of the best studied plasmids is pME2001 of *M.thermoautotrophicum* Marburg. This plasmid which is present in about 20 copies per host chromosome (102), turned out to be very stable since several attempts failed to cure strain Marburg from pME2001 (102). The dispensability of pME2001 was, however, demonstrated by Leisinger and coworkers who reported the isolation of a pME2001-free derivative of strain Marburg after treatment with the mutagen N-methyl-N'-nitro-nitrosoguanidine (70, 92). The nucleotide sequence of pME2001 comprises 4439 bp (7) and contains six open reading frames, designated ORF A through ORF F, which could code for proteins composed of more than 100 amino acids (Figure 2A). Except for ORF C, all ORFs are preceded by a potential ribosome binding site (10). The head-to-tail-organization of ORF A-D may suggest a coordinate transcription and translation of the four ORFs. Although no sequences with significant similarity to the *boxA/boxB* motif of an archaeal consensus promoter signal (44, 48) could be localized upstream from ORF A, the identification of a 611-bp-transcript of pME2001 by Meile and coworkers (91) which covers the potential promoter regions of the ORF A and ORF F, respectively (Figure 2A), may support cotranscription of ORF A-D. The small size of the transcript may be explained as the result of posttranscriptional processing or nucleolytic degradation of a primary transcript (91). A sequence stretch with excellent similarity to the archaeal consensus promoter signal is located approximately 170 bp upstream of ORF F. Possibly, the expression of the ORF F protein is regulated since translation of the ORF F mRNA might be blocked by the mapped 611-bp-RNA which is transcribed from the opposite DNA-strand and would therefore serve as anti-sense RNA (Figure 2A).

No significant similarity was found between the deduced proteins of pME2001 and proteins stored in the database. Screening of the pME2001 proteins for the presence of certain amino acids motifs has shown, however, that the ORF B protein contains a helix-turn-helix (HTH) motif which might enable the ORF B protein to bind to DNA (Nölling, unpublished results). Moreover, the HTH motif of the ORF B protein was found to be similar with the HTH motif of the transcription activator cII from phage λ (51). The identical residues shared by both HTH motifs were in particular located at those positions of the λ cII protein that are believed to be involved in recognition of a specific DNA sequence (51) (Figure 2B). This may suggest that the ORF B protein recognizes the same DNA target as the λ cII transcription activator, i.e. two TTGC tetranucleotides repeats spaced by six nucleotides: TTGC-N_6-TTGC (51). We could identify a region located at the 3'-end of ORF D of plasmid pME2001 that showed homology to the λ cII target-sequence. This pME2001-region, designated TA-region, is 193

bp in size and contains three sequence stretches (TA1-TA3) with the same or similar tetranucleotide repeats separated by six nucleotides as found for the λ cII-target (Figure 2C). An additional remarkable feature of TA1-TA3 sequences is the fact that those tetranucleotide repeats in turn separate several runs of four to six thymidine nucleotides which may induce bending of the DNA (78) (Figure 2C). Such bent DNA regions are known to be important for various molecular-biological processes such as for instance replication and transcription (43). It seems therefore possible that the putative DNA binding ORF B protein and the TA-region play a role in replication of plasmid pME2001.

Several pME2001-based vectors have been constructed that may be used for replication in *Bacteria* and *Eucarya* (89, 90). However, besides a low-level natural transformation system for *M.thermoautotrophicum* Marburg (162), no efficient reintroduction of DNA into this methanogenic host has been reported so far. The development of a transformation system is therefore the bottle neck for examination of *in vitro* modified DNA in *Methanobacterium* strains.

Promising results in establishing a transformation have been reported for the mesophilic *Methanococcus voltae* PS. Based on its low-level natural competence (4), transformation of this organism with a vector that was able to integrate in the chromosome of the recipient, was demonstrated by Gernhardt and coworkers (36). The integration vector used was equipped with an expression cassette comprising the *puc* gene from *Streptomyces alboniger* (83) as selective marker which conferred puromycin resistance to its methanogenic host. Similar results were obtained with *Methanococcus maripaludis* used as recipient (128). Increased transformation rates of auxotrophic *M.voltae* PS with wild type chromosomal DNA could be achieved by electroporation as alternative transformation method (95). These results suggest that an efficient host-vector system for *Methanococcus* could be based on an improved electroporation protocol in combination with an autonomously replicating vector that carries the puromycin-resistance cassette (36) as selective marker. The vector may be derived from a plasmid of a related species, such as pURB500 isolated from a marine *Methanococcus* species (160).

Viruses and Virus-like Particles

Similar as representatives of the eucaryal and bacterial domain also those belonging to the archaeal domain are subject to viral infection. Table 2 lists the viruses and virus-like particles (VPLs: nucleic acid containing particles whose infectivity have not been proven) that have been described for representatives of the three major lineages of the archaeal domain, the extreme thermophilic, sulfur-dependent, the extreme halophilic and the methanogenic *Archaea* (for review: 172). All archaeal viruses and VLPs, isolated so far, contain a double-stranded DNA genome although this does not exclude the possibility of the existence of archaeal viruses with single-stranded DNA or RNA genomes. With the exception of the VLP isolated from *Methanococcus voltae* A3 (161), viruses of the extreme halophilic and methanogenic branch were similar in morphology to those of the bacterial domain composed of a polyhedral head and a tail. By contrast, unusual lemon- or rod-shaped virus structures are common to the viruses and VPLs that infect extreme thermophilic, sulfur-dependent *Archaea*. In addition, the genomes of the latter viruses and VLPs are only small in size compared to those found for the viruses of halophiles and methanogens (Table 2).

The temperate virus SSV1 (formerly SAV1) of the extremely thermophilic, sulfur-

Table 2. Archaeal viruses and virus-like particles

Name	Host	Virion morphology	DNA topology	Genome size (kb)	Type	References
SSV1	*Sulfolobus solfataricus*	lemon	ccc[a]	15.5	temperate	(86, 109, 130, 168)
VPL-B6[b]	*Sulfolobus sp. B6*	polyhedric	n.d.[c]	n.d.	temperate	(172)
DAV 1[d]	*Desulfurococcus ambivalens*	lemon	ccc	7.7	temperate	(167, 168, 172)
CWP	*Pyrococcus woesei*	polyhedric	n.d.	n.d.	n.d.	(170, 172)
TTV1	*Thermoproteus tenax* Kra1	rod	linear	15.9	temperate[e]	(63, 98, 172)
TTV2	*T.tenax* Kra1	flexible rod	linear	16	temperate	(63, 172)
TTV3	*T.tenax* Kra1	flexible rod	linear	27	n.d.	(63, 172)
TTV3	*T.tenax* Kra1	rod	linear	17	lytic	(63, 172)
ΦH	*Halobacterium halobium*[f]	complex[g]	linear	59	temperate	(38, 131, 135, 172)
ΦN	*H.halobium*	complex	linear	56	temperate	(151)
Hs1	*H.salinarium*	complex	linear	n.d.	temperate[e]	(147, 148, 149)
HsL						
Hh1	*Halobacterium* sp.	complex	linear	37.2	temperate[e]	(110)
Hh3	*Halobacterium* sp.	complex	linear	29.4	temperate[e]	(110)
Ja1	*H.halobium*	complex	linear	230	n.d.	(152)
S45	*Halobacterium* sp.	complex	linear	n.d.	n.d.	(21)
VLP(A3)	*Methanococcus voltae* A3	lemon	ccc	23	temperate	(161)
PG	*Methanobrevibacter smithii* G	complex	n.d.	50	lytic	(3)
PMS1	*M.smithii*	n.d.	n.d.	35	lytic	(76)
ψM1	*Methanobacterium thermoautotrophicum* Marburg	complex	linear	27.1	lytic	(65, 92)
ΦF1	*Methanobacterium* sp.	complex	linear	85	lytic	(108)
ΦF3	*M.thermoformicicum* FF3	complex	ccc	36	lytic	(108)

[a] Covalently closed circular
[b] The VLP-B6 genome is suggested to correspond with plasmid pBG (172).
[c] Not determined.
[d] The DAV 1 genome is identical with plasmid pSL10 (172).
[e] So-called 'carrier state' defined as an equilibrium between virus and cell multiplication. A shift of this equilibrium results in either loss of the phage genome or induction of the lytic circle (172).
[f] Several strains of *H.halobium* belong to the species *H.salinarium* (146).
[g] Consisting of an icosahedric head and a tail.

dependent *Sulfolobus shibatae* (86) is one of the archaeal viruses studied in detail on the molecular level. Initially described as VLP, SSV1 was recently shown to infect the closely related *S.solfataricus* and is thus a true virus (130). The virus SSV1 contains a circular, positively supercoiled (97) double-stranded DNA genome with a size of 15.5 kb. Its complete nucleotide sequence has been determined (109) and the transcriptionally active regions have been mapped (123). Upon lysogenisation, the SSV1 genome is either maintained as extrachromosomal, autonomous replicating DNA (formerly termed pSB12) or is integrated into the chromosome (164). Integration occurs site-specifically in both *S.shibatae* and *S.solfataricus* within a putative $tRNA^{Arg}$ gene (124, 130). UV-irradiation induces amplification and packaging of the SSV1 plasmid form and the release of phage particles without lysis of the host (86, 130, 164). Since efficient transfection (10^6 transfectants per μg

DNA) of *S.solfataricus* with SSV1 DNA by electroporation has been demonstrated (130), the well-studied SSV1 genome may be instrumental in developing a cloning vector for *Sulfolobus*.

Four rod-shaped or filamentous viruses, designated TTV1 through TTV4, have been described for *Thermoproteus tenax* Kra1 (63). Although these viruses exhibit a similar morphology they clearly differ in several aspects. TTV2 is a temperate virus, whereas TTV4 is lytic, while cells carrying TTV1 may either lyse or lose the virus (172). The viral genomes consist of linear, double-stranded DNA molecules of different size that showed no homology with each other (169). Moreover, only limited similarities have been observed in the composition of the virus envelope (169). TTV1 has been studied in most detail. The TTV1 DNA is covered by two DNA-binding proteins contained within the proteinaceous inner envelope that is surrounded by a asymmetric membrane (172). The genes encoding the two DNA-binding and the three major structural proteins of TTV1 have been analyzed (98, 172). Six genomic variants of TTV1 could be isolated that differed from each other in a distinct region by deletions and insertions (99). Since a phage-resistant strain of *T.tenax* contains part of this region integrated into its chromosome, it may specify immunity functions (172). Sequence analysis of the region revealed an unusual 465-bp ORF that could code for a protein with highly repeated amino acid stretches (169).

Three VPLs have been isolated from sulfur-dependent *Archaea*, CWP from the hyperthermophilic *Pyrococcus woesei*, DAV1 from *Desulfurolobus ambivalens* and a VPL from *Sulfolobus* strain B6. Since no nucleic acids could be isolated from CWP particles the nature of their genome remains unknown (170). Similarly, no data are available that clearly prove the association of the VPLs produced by *Sulfolobus* B6 with the genetic variable 35-kb plasmid pBG detected in the host (143, 172). In contrast, nucleic acids isolated from DAV1 particles seem to correspond with the 7.7-kb plasmid pLS10 of *Desulfurolobus ambivalens*, an extremely thermophilic, facultatively anaerobic archaeon (167, 168, 172). Interestingly, pLS10 shares some similarity with SSV1 DNA (172). Since anaerobic growth of the host is accompanied with amplification of plasmid pSL10 and the production of virus-like particles, pLS10 probably represents the lysogenic form of the virus DAV1 (169).

Phages isolated from the halophilic branch of the *Archaea* differ considerably in their genome sizes ranging from 30 kb for phage Hh3 (110) up to 230 kb reported for phage Ja1 (152). One of the most extensively studied halophilic virus-host systems is that of phage ΦH and its host *Halobacterium halobium* R_1 (131, 172). The linear 59-kb genome of the temperate phage ΦH, whose transcription units have been mapped (172), is circularized upon lysogenization and maintained as extrachromosomally replicating prophage (135). Similar to the genome of its host, a high degree of genetic variability has been observed for ΦH that is due to the presence of certain IS-elements (132). One of these elements, ISH1.8 (134; see below), has been found to flank a central 12-kb fragment, called L-region, whose inversion or looping out is mediated by ISH1.8 (132, 134). The L-region contains genes for both immunity and the lytic cycle and obviously specifies all functions necessary for replication that enables the L-plasmid to be stably maintained in its host (see also plasmid section). Efficient transfection (approximately 10^7 transfectants per μg DNA) of *H.halobium* with ΦH DNA has also been reported based on a protocol similar to those used for plasmid transformation of halobacteria (15; see plasmid section). DNA modification has been described for another phage from *H.halobium*, ΦN, whose linear, 56-kb genome contains 5-methyl-cytosine (151).

A variety of viruses and VLPs have been described for methanogenic bacteria (for review:

172; Table 2). VLPs were found to accumulate during growth of *Methanococcus voltae* A3 (161). Similar to the prophage state of SSV1 in *Sulfolobus shibatae*, the genome of the *M.voltae* A3-derived VLPs which is a circular 23-kb DNA molecule, termed pURB600, could be detected either as free plasmid or integrated into the chromosome of its host. In addition, pURB600-related sequences have been identified in the chromosome of *M.voltae* PS which, however, did not produce VPLs (161). Plasmid pMP1 of *Methanolobus vulcani* (see plasmid section) may also represent a prophage since it has been found integrated into the host chromosome (155). The isolation of two lytic phage particles infecting *Methanobrevibacter smithii*, the broad host range phage PG (3) and phage PMS1 (76), with genome sizes of 50 and 35 kb, respectively, has been reported. The model virus-host system of methanogens that has been studied in some detail is that of the lytic phage ψM1 and its host *Methanobacterium thermoautotrophicum* Marburg (92). The genome of phage ψM1 is a linear, circularly permutated 27.1-kb DNA molecule that displays a terminal redundancy of approximately 3 kb (65). Packaging of the phage DNA is initiated from a concatemeric precursor molecule at a mapped *pac* site by the 'headful packaging' model (65). Interestingly, about 15% of the ψM1 phage particles was found to contain hexamers of the cryptic 4.4-kb plasmid pME2001 present in the strain Marburg (88, 92). This relaxed packaging machinery of ψM1 may be responsible for the recently demonstrated general transduction of chromosomal markers mediated by ψM1 (93). A 16-kb chromosomal region with homology to ψM1 DNA was identified in *M.wolfei* suggesting that this organism contains a deletion derivative of ψM1 (92). This may explain the similarity of the endopeptidases capable to hydrolyse pseudomurein-based cell walls that have been purified from lysates of autolysed *M.wolfei* cultures (71) and lysates produced by *M.thermoautotrophicum* Marburg after ψM1 infection (142).

Insertion sequences

Most archaeal IS elements have been isolated from halophilic organisms, whereas only one IS elements have been reported for methanogens (Table 3). So far no information is available about such mobile elements in extreme thermophilic, sulfur-dependent *Archaea*.

The biology of IS elements has intensively been studied in the halophilic model organism *Halobacterium halobium*. *H.halobium* displays a considerable genetic instability that results in high spontaneous mutation rates of phenotypic markers such as gas vacuole formation and synthesis of the purple membrane protein bacterio-opsin which are lost with frequencies of 10^{-2} and 10^{-4}, respectively (113, 153). Most of these mutants were shown to result from genomic rearrangements such as insertions, deletions and recombination events induced by various IS-elements (22, 24, 25, 115, 116, 138). As indicated by further studies, the majority of these transposable elements and other repeated sequences are clustered within the relatively AT-rich satellite DNA fraction (GC-content 58%) consisting of plasmid DNA and chromosomal 'islands' that comprise about 30% of the genome of extreme halophilic organisms such as *H.halobium* (114, 118).

IS elements from halophiles (ISH elements; for review: 13) range in size from 520 bp for ISH2 up to 3000 bp for ISH24 (Table 3) and resemble in their structure 'typical' insertion sequences described for the bacterial domain (35), i.e., they contain terminal inverted repeats and are flanked by short direct repeats as the result of target site duplication (Table 3). ISH1.8 is an exception in that it lacks terminal inverted repeats (134). IS elements are defined as autonomously moving elements, i.e. they mediate their own transposition, which

Table 3. Archaeal insertion elements

Element	Host[a]	Element length (bp)	ORF[b] length (bp)	Copy number (bp)	IR[c] length (bp)	DR[d] length	Indentical copies[e]	References
ISH1	*Halobacterium halobium*[f]	1118	810	2	7	8	yes	(138)
ISH1.8	*H.halobium*	1895	max. 672[g]	2	—	5	n.d.[h]	(134)
ISH2	*H.halobium*	520	—	8-10	19	10-11, 20	yes	(22)
ISH11	*H.halobium*	1068	1002	n.d.	15	7	n.d.	(81)
ISH23/50	*H.halobium*	996	819, 366	n.d.	19	9/8	n.d.	(115, 116, 163)
ISH24	*H.halobium*	3000	n.d.	2	14	7	n.d.	(115, 116)
ISH26	*H.halobium*	1384	441, 741	5	15	9, 11	no	(29)
ISH27	*H.halobium*	1398	1167	>2	16	5	no	(117, 120)
ISH28	*H.halobium*	932	0	>2	16	8	n.d.	(117, 121)
ISH S1	*H.halobium*	1449	1179, 366	n.d.	26	0	n.d.	(175)
ISH51	*H.volcanii*	1375	1146	20-30	16	3	no	(52, 120)
ISM1	*Methanobrevibacter smithii*	1381	1203	8-10	29	8	n.d.	(46)
FR-I	*Methanobacterium thermoformicicum*	1501	750[i]	4-5	—	6	yes	(107)

[a] Original host.
[b] Open reading frames with more than 300 bp.
[c] Terminal located inverted repeats.
[d] Direct repeats flanking the element.
[e] Nucleotide sequence idendity of different copies.
[f] Several strains of *H.halobium* belong to the species *H.salinarium* (146).
[g] ISH1.8 contains six ORFs with a size of more than 300 bp.
[h] not determined.
[i] FR-I exists in two structural variants one of which contains a 750 bp ORF.

is catalyzed by an element-encoded transposase activity. Most of the ISH elements listed in Table 3 indeed contain at least one ORF that may encode a transposase-like protein. Some of the ISH elements, however, lack appropriate ORFs (Table 3) and probably represent non-autonomous IS elements that may transpose by transposase functions provided in *trans* by other, autonomous ISH elements. This may be the case for ISH51 from *Haloferax volcanii*. From three ISH51 elements that have been analyzed on the sequence level, only one was found to contain a large ORF (52, 120). Interestingly, ISH27 from *H.halobium* and ISH51 constitute a family of related insertion sequences that show identical terminal inverted repeats and considerable similarity on nucleotide sequence level of the entire elements (120). In the case of ISH2 which has an only low coding capacity (22), the existence of a copy with an intact function necessary for transposition seem to be unlikely since all so far analyzed ISH2 copies were identical on sequence level (119). However, the terminal inverted repeats of ISH2 are considerable similar to those found for ISH26 whereas the internal sequences of both elements lack homology (29). Since terminal inverted repeats play an important role in transposition of IS elements it was suggested that ISH26 mediates transposition of ISH2 (29).

Among methanogens only one transposable element, ISM1 from *Methanobrevibacter smithii*, has been characterized previously (46; Table 3). ISM1, whose movement within the host genome has been demonstrated, shows all features of a 'typical' insertion elements and contains a large ORF that probably specifies the transposase function (46).

OUTLINE OF THE THESIS

The limited knowledge about plasmid biology of methanogens on the one hand and the need for developing host-vector system in this branch of *Archaea* on the other hand, were the principal motives to initiate the present thesis. Accordingly, the project had a basic character and was aimed at analyzing newly isolated plasmid DNA from methanogenic bacteria on molecular level with emphasis on those plasmids harbored by thermophilic strains of the genus *Methanobacterium*. After a brief summary of the state of art (Chapter 1), a phylogenetic analysis of different thermophilic *Methanobacterium* strains including several new isolates is described in Chapter 2. The obtained results from sequence comparison of PCR-amplified 16S rRNA genes strongly suggested a reclassification of the species *M.thermoformicicum* and *M.thermoautotrophicum*. The following Chapters 3 and 4 deal with the characterization of two types of mobile elements, namely phages and plasmids. The isolation and characterization of two novel phages, ΦF1 and ΦF3, capable of infecting several thermophilic *Methanobacterium* strains is reported in Chapter 3. Phage ΦF1 whose linear genome is approximately 80 kb in size, has a broad host range and could be propagated on *M.thermoformicicum* Z-245, FTF, FF1, FF3 and CSM3 as well as *M.thermoautotrophicum* ΔH. In contrast, phage ΦF3 is specific for *M.thermoformicicum* FF3. Chapter 4 describes the isolation and characterization of a family of related plasmids from the thermophilic archaeon *M.thermoformicicum* and initial approaches in classification of *M.thermoformicicum* strains by genetic fingerprinting. Two of the plasmids, pFV1 from strain THF and pFZ1 from strain Z-245, were analysed in detail. Comparison of the complete nucleotide sequences of pFV1 and pFZ1 allowed an outline of the genetic organization of the plasmid genomes consisting of conserved regions which are interspersed with accessory elements (Chapter 5). Moreover, evidence is provided that plasmid pFV1 encodes a putative DNA mismatch repair which is functionally related to the pFV1-borne restriction-modification (R/M) system (Chapter 5). One of the accessory elements, termed FR-I, was present in the genomes of strains of *M.thermoformicicum* and *M.thermoautotrophicum* (Chapter 6). Although transposition of FR-I has not been demonstrated, several characteristics of FR-I support the hypothesis that it represents a new type of archaeal insertion elements. In addition, pFV1 and pFZ1 were found to contain an accessory element encoding components of different R/M systems, whose detailed analysis is discussed in Chapter 7 and 8, respectively. Finally, a summary and concluding remarks are given in Chapter 9.

REFERENCES

1. **Amábile-Cuevas, C. F., and M. E. Chicurel.** (1992). Bacterial plasmids and gene flux. Cell **70**, 189-199.
2. **Berry, D. F., A. J. Francis, and J. M. Bolleg.** (1987). Microbial metabolism of homocyclic and heterocyclic compounds under anaerobic conditions. Microbiol. Rev. **51**, 43-59.
3. **Bertani, G., and L. Baresi.** (1986). Looking for a gene transfer mechanism in methanogenic bacteria. System. Appl. Microbiol. **7**, 398.
4. **Bertani, G., and L. Baresi.** (1987). Genetic transformation in the methanogen *Methanococcus voltae*. J. Bacteriol. **169**, 2730-2738.
5. **Biavati, B., B. Sgorbati, and D. Palenzona.** (1992). Plasmid DNA from methanogenic bacteria. Curr. Microbiol. **24**, 285-287.

6. **Blaseio, U., and F. Pfeifer.** (1990). Transformation of *Halobacterium halobium*: development of vectors and investigation of gas vesicle synthesis. Proc. Natl. Acad. Sci. USA **87**, 6772-6776.
7. **Bokranz, M., A. Klein, and L. Meile.** (1990). Complete nucleotide sequence of plasmid pME2001 of *Methanobacterium thermoautotrophicum* (Marburg). Nucleic Acids Res. **18**, 363.
8. **Bouthier de la Tour, C., C. Portemer, M. Nadal, K. O. Stetter, P. Forterre, and M. Duguet.** (1990). Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J. Bacteriol. **172**, 6803-6808.
9. **Bouthier de la Tour, C., C. Portemer, R. Huber, P. Forterre, and M. Duguet.** (1991). Reverse gyrase in thermophilic eubacteria. J. Bacteriol. **173**, 3921-3923.
10. **Brown, J. W., C. J. Daniels, and J. N. Reeve.** (1989). Gene structure, organization and expression in archaebacteria. Crit. Rev. Microbiol. **16**, 287-337.
11. **Charbonnier, G. Erauso, T. Barbeyron, D. Prieur, and P. Forterre.** (1992). Evidence that a plasmid from a hyperthermophilic archaebacterium is relaxed at physiological temperatures. J. Bacteriol. **174**, 6103-6108.
12. **Charlebois, R. L., W. L. Lam, S. W. Cline, and W. F. Doolittle.** (1987). Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. Proc. Natl. Acad. Sci. USA **84**, 8530-8534.
13. **Charlebois, R. L., and W. F. Doolittle.** (1989). Transposable elements and genome structure in Halobacteria. In: Mobile DNA, p. 297-308 (eds. D. G. Berg and M. M. Howe), American Society for Microbiology, Washington DC.
14. **Charlebois, R. L., L. C. Schalkwyk, J. D. Hofman, and W. F. Doolittle.** (1991). Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol. **222**, 509-524.
15. **Cline, S. W., and W. F. Doolittle.** (1987). Efficient transfection of the archaebacterium *Halobacterium halobium*. J. Bacteriol. **169**, 1341-1344.
16. **Cline, S. W., W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, and W. F. Doolittle.** (1989). Transformation methods for halophilic bacteria. Can. J. Microbiol. **35**, 148-152.
17. **Cline, S. W., L. C. Schalkwyk, and W. F. Doolittle.** (1989). Transformation of the archaebacterium *Halobacterium volcanii* with genomic DNA. J. Bacteriol. **171**, 4987-4991.
18. **Cline, S. W., and W. F. Doolittle.** (1992). Transformation of members of the genus *Haloarcula* with shuttle vectors based on *Halobacterium halobium* and *Haloferax volcanii* plasmid replicons. J. Bacteriol. **174**, 1076-1080.
19. **Cowan, D. A.** (1992). Biotechnology of the Archaea. Trends Biotechnol. **10**, 315-323.
20. **Creti, R., F. Citarella, O. Tiboni, A. Sanangelantoni, P. Palm, and P. Cammarano.** (1990). Nucleotide sequence of a DNA region comprising the gene for elongation factor 1α (EF-1α) from the ultrathermophilic archaeote *Pyrococcus woesei*: phylogenetic implications. J. Mol. Evol. **33**, 332-342.
21. **Daniels, L. L., and A. C. Wais.** (1984). Restriction and modification of halophage S45 in *Halobacterium*. Curr. Microbiol. **10**, 133-136.
22. **DasSarma, S., U. L. RajBhandary, and H. G. Khorana.** (1983). High frequency spontaneous mutations in the bacterio-opsin gene of *Halobacterium halobium* are mediated by transposable elements. Proc. Natl. Acad. Sci. USA **80**, 2201-2205.
23. **DasSarma, S., T. Darnerval, J. G. Jones, and N. Tandeau de Marsac.** (1987). A plasmid-encoded gas vesicle protein gene in a halophilic archaebacterium. Mol. Microbiol. **1**, 365-370.
24. **DasSarma, S., J. T. Halladay, J. G. Jones, J. W. Donovan, P. J. Giannasca, and N. Tandeau de Marsac.** (1988). High frequency mutations in a plasmid-encoded gas vesicle gene in *Halobacterium halobium*. Proc. Natl. Acad. Sci. USA **85**, 6861-6865.
25. **DasSarma, S.** (1989). Mechanisms of genetic variability in *Halobacterium halobium*: t h e purple membrane and gas vesicle mutations. Can. J. Microbiol. **35**, 65-72.
26. **de Rosa, M., A. Cambacorta, and A. Gliozzi.** (1986). Structure, biosynthesis, and

physiochemical properties of archaebacterial lipids. Microbiol. Rev. **50**, 70-80.
27. **DiMarco, A. A., T. A. Bobik, and R. S. Wolfe.** (1990). Unusual coenzymes of methanogenesis. Annu. Rev. Biochem. **59**, 355-394.
28. **Ebert, K., W. Goebel, and F. Pfeifer.** (1984). Homologies between heterogeneous extrachromosomal DNA populations of *Halobacterium halobium* and four new isolates. Mol. Gen. Genet. **194**, 91-97.
29. **Ebert, K., C. Hanke, H. Delius, W. Goebel, and F. Pfeifer.** (1987). A new insertion element, ISH26, from *Halobacterium halobium*. Mol. Gen. Genet. **206**, 81-87.
30. **Edmonds, C. G., P. F. Crain, R. Gupta, T. Hashizume, C. H. Hocart, J. A. Kowalak, S. C. Pomerantz, K. O. Stetter, and J. A. McCloskey.** (1991). Posttranscriptional modification of tRNA in thermophilic Archaea (archaebacteria). J. Bacteriol. **173**, 3138-3148.
31. **El Hassouni, M., B. Henrissat, M. Chippaux, and F. Barras.** (1992). Nucleotide sequence of the *arb* genes, which control β-glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new β-glycohydrolase family including enzymes from eubacteria, archaebacteria, and humans. J. Bacteriol. **174**, 765-777.
32. **Erauso, G., F. Charbonnier, T. Barbeyron, P. Forterre, and D. Prieur.** (1992). Preliminary characterization of a hyperthermophilic archaebacterium with a plasmid, isolated from a North Fiji basin hydrothermal vent. C. R. Acad. Sci. Paris **314**, 387-393.
33. **Fabry, S., P. Heppner, W. Dietmaier, and R. Hensel.** (1990). Cloning and sequencing the gene encoding 3-phosphoglycerate kinase from mesophilic *Methanobacterium bryantii* and the thermophilic *Methanothermus fervidus*. Gene **91**, 19-95.
34. **Forterre, P., F. Charbonnier, E. Marguet, F. Haper, and G. Henckes.** (1992). Chromosome structure and DNA topology in extreme thermophilic archaebacteria. Biochemical Soc. Symp. **58**, in press.
35. **Galas, D. J., and M. Chandler.** (1989). Bacterial insertion sequences. In: Mobile DNA, p. 109-162 (eds. D. E. Berg and M. M. Howe), American Society for Microbiology, Washington DC.
36. **Gernhardt, P., O. Possot, M. Foglino, L. Sibold, and A. Klein.** (1990). Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. Mol. Gen. Genet. **221**, 273-279.
37. **Gogarten, J. P., H. Kibak, P. Dittrich, L. Taiz, E. J. bowman, B. J. Bowman, M. F. Manolson, R. J. Poole, T. Date, T. Oshima, J. Konoshi, K. Denda, and M. Yoshida.** (1989). Evolution of the vacuolar H^+-ATPases: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA **86**, 6661-6665.
38. **Gropp, F., B. Grampp, P. Stolt, P. Palm, and W. Zillig.** (1992). The immunity-conferring plasmid pΦHL from the *Halobacterium salinarium* phage ΦH: nucleotide sequence and transcription. Virology **190**, 45-54.
39. **Gruss, A., and S. D. Ehrlich.** (1989). The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol. Rev. **53**, 231-241.
40. **Gutiérrez, M. C., M. T. García, A. Ventosa, J. J. Nieto, and F. Ruiz-Berraquero.** (1986). Occurrence of megaplasmids in halobacteria. J. Appl. Bacteriol. **61**, 67-71.
41. **Hackett, N. R., and S. DasSarma.** (1989). Characterization of the small endogenous plasmid of *Halobacterium* strain SB3 and its use in transformation of *H.halobium*. Can. J. Microbiol. **35**, 86-91.
42. **Hackett, N. R., M. P. Krebs, S. DasSarma, W. Goebel, U. L. RajBhandary, and H. G. Khorana.** (1990). Nucleotide sequence of a high copy number plasmid from *Halobacterium* strain GRB. Nucleic Acids. Res. **18**, 3408.
43. **Hagerman, P. J.** (1990). Sequence-directed curvature of DNA. Annu. Rev. Biochem. **59**, 755-781.
44. **Hain, J., W.-D. Reiter, U. Hüdepohl, and W. Zillig.** (1992). Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Res. **20**, 5423-5428.

45. **Hall, M. J., and N. R. Hackett.** (1989). DNA sequence of a small plasmid from *Halobacterium* strain GN101. Nucleic Acids Res. **17**, 10501.
46. **Hamilton, P. T., and J. N. Reeve.** (1985). Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. Mol. Gen. Genet. **200**, 47-59.
47. **Hasegawa, M., N. Iwabe, Y. Mukohata, and T. Miyata.** (1990). Close evolutionary relatedness of archaebacteria, *Methanococcus* and *Halobacterium*, to eukaryotes demonstrated by composite phylogenetic trees of elongation factors EF-Tu and EF-G. Eocyte tree is unlikely. Jpn. J. Genet. **65**, 109-114.
48. **Hausner, W., G. Frey, and M. Thomm.** (1991). Control regions of an archaeal gene: a TATA box and an initiator element promote cell-free transcription of the $tRNA^{Val}$ gene of *Methanococcus voltae*. J. Mol. Biol. **222**, 495-508.
49. **Heinemann, J. A.** (1991). Genetics of gene transfer between species. Trends in Genetics **7**, 181-185.
50. **Hensel, R., P. Zwickle, S. Fabry, J. Lang, and P. Palm.** (1989). Sequence comparison of the glyceraldehyde-3-phosphate dehydrogenase from the three urkingdoms: evolutionary implications. Can. J. Microbiol. **35**, 81-85.
51. **Ho, Y. S., and M. Rosenberg.** (1988). Structure and function of the transcription activator protein cII and its regulatory signals. In: The Bacteriophages, Vol. 2, p. 725-756 (eds. R. Calendar), Plenum Press, New York-London.
52. **Hofman, J. D., L. C. Schalkwyk, and W. F. Doolittle.** (1986). ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaebacterium *Halobacterium volcanii*. Nucleic Acids Res. **14**, 6983-7000.
53. **Holliger, C., S. W. M. Kengen, G. Schraa, A. J. M. Stams, and A. J. B. Zehnder.** (1992). Methyl-coenzyme M reductase of *Methanobacterium thermoautotrophicum* ΔH catalyses the reductive dechlorination of 1,2-dichloroethane to ethylene and chloroethane. J. Bacteriol. **174**, 4435-4443.
54. **Holmes, M. L., and M. L. Dyall-Smith.** (1990) A plasmid vector with a selectable marker for halophilic archaebacteria. J. Bacteriol. **172**, 756-761.
55. **Holmes, M. L., and M. L. Dyall-Smith.** (1991). Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. J. Bacteriol. **173**, 642-648.
56. **Holmes, M. L., S. D. Nuttall, and M. L. Dyall-Smith.** (1991). Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. J. Bacteriol. **173**, 3807-3813.
57. **Honka, E., S. Fabry, T. Nierman, P. Palm, and R. Hensel.** (1990). Properties and primary structure of the L-malate dehydrogenase from the extremely thermophilic archaebacterium *Methanothermus fervidus*. Eur. J. Biochem. **188**, 623-632.
58. **Horne, M., C. Englert, and F. Pfeifer.** (1988). Two genes encoding gas vacuole proteins in *Halobacterium halobium*. Mol. Gen. Genet. **213**, 459-464.
59. **Horne, M., C. Englert, C. Wimmer, and F. Pfeifer.** (1991). A DNA region of 9 kbp contains all genes necessary for gas vesicle synthesis in halophilic archaebacteria. Mol. Microbiol. **5**, 1159-1174.
60. **Ihara, K., and Y. Mukohata.** (1991). The ATP synthase of *Halobacterium salinarium (halobium)* is an archaebacterial type as revealed from the amino acid sequences of its two major subunits. Arch. Biochem. Biophys. **286**, 111-116.
61. **Ilyina, T. V., and E. V. Koonin.** (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. **20**, 3279-3285.
62. **Iwabe, N., K.-I. Kuma, M. Hasegawa, S. Osawa, and T. Miyata.** (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA **86**, 9355-9359.

63. **Janeković, D., S. Wunderl, I. Holz, W. Zillig, A. Gierl. and H. Neumann.** (1983). TTV1, TTV2 and TTV3, a family of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium *Thermoproteus tenax*. Mol. Gen. Genet. **192**, 39-45.
64. **Jones, J. G., N. R. Hackett, J. T. Halladay, D. J. Scothorn, C. Yang, W. Ng, and S. DasSarma.** (1989). Analysis of insertion mutants reveals two new genes in the pNRC100 gas vesicle gene cluster of *Halobacterium halobium*. Nucleic Acids Res. **17**, 7785-7793.
65. **Jordan, M., L. Meile, and T. Leisinger.** (1989). Organization of *Methanobacterium thermoautotrophicum* bacteriophage ψM1. Mol. Gen. Genet. **220**, 161-164.
66. **Kaine, B. P.** (1990). Structure of the archaebacterial 7S RNA molecule. Mol. Gen. Genet. **221**, 315-321.
67. **Kamekura, M., Y. Seno, M. L. Holmes, and M. L. Dyall-Smith.** (1992). Molecular cloning and sequencing of the gene for a halophilic alkaline serine protease (halolysin) from an unidentified halophilic Archaea strain (172P1) and expression of the gene in *Haloferax volcanii*. J. Bacteriol. **174**, 736-742.
68. **Keltjens, J. T., and C. van der Drift.** (1986). Electron transfer reactions in methanogens. FEMS Microbiol. Rev. **39**, 259-303.
69. **Ken, R., and N. R. Hackett.** (1991). *Halobacterium halobium* strains lysogenic for phage ΦH contain a protein resembling coliphage repressors. J. Bacteriol. **173**, 955-960.
70. **Kiener, A., T. Rechsteiner, and T. Leisinger.** (1986). Mutation to pseudomonic acid resistance of *Methanobacterium thermoautotrophicum* leads to an altered isoleucyl-tRNA synthetase. FEMS Lett. **33**, 15-18.
71. **Kiener, A., H. König, and T. Leisinger.** (1987). Purification and use of *Methanobacterium wolfei* pseudomurein endopeptidase for lysis of *Methanobacterium thermoautotrophicum*. J. Bacteriol. **169**, 1010-1016.
72. **Kimura, M.** (1983). The neutral theory of molecular evolution. Cambridge University Press, New York.
73. **Kjems, J., and R. A. Garrett.** (1985). An intron in the 23S ribosomal RNA gene of the archaebacterium *Desulfurococcus mobilis*. Nature (London) **318**, 675-677.
74. **Klenk, H.-P., P. Palm, F. Lottspeich, and W. Zillig.** (1992). Component of the DNA-dependent RNA polymerases of *Archaea* is homologous to a subunit shared by the three eucaryal nuclear RNA polymerases. Proc. Natl. Acad. Sci. USA **89**, 407-410.
75. **Kletzin, A.** (1992). Molecular characterization of a DNA ligase gene of the extremely thermophilic archaeon *Desulfurolobus ambivalens* shows close phylogenetic relationship to eukaryotic ligases. Nucleic Acids Res. **20**, 5389-5396.
76. **Knox, M. R., and J. E. Harris.** (1986). Abstr. XIV Internat. Congr. Microbiol., PG3-8, p. 240.
77. **Konisky, J.** (1989). Methanogens for biotechnology: application of genetics and molecular biology. Trends Biotechnol. **7**, 88-92.
78. **Koo, H.-S., Wu, H.-M., and D. M. Crothers.** (1986). DNA bending at adenine · thymidine tracts. Nature (London) **320**, 501-506.
79. **Kornberg, A., and T. Baker.** (1992). DNA Replication. 2nd ed., W. H. Freeman and Company, New York.
80. **Krawiec, S., and M. Riley.** (1990). Organization of the bacterial chromosome. Microbiol. Rev. **54**, 502-539.
81. **Krebs, M. P., U. L. RajBhandary, and H. G. Khorana.** (1990). Nucleotide sequence of ISH*11*, a new *Halobacterium halobium* insertion element isolated form the plasmid pGRB1. Nucleic Acids Res. **18**, 6699.
82. **Krebs, M. P., T. Hauss, M. P. Heyn, U. L. RajBhandary, and H. G. Khorana.** (1991). Expression of the bacterioopsin gene in *Halobacterium halobium* using a multicopy plasmid. Proc. Natl. Acad. Sci. USA **88**, 859-863.
83. **Lacalle, R. A., D. Pulido, J. Vara, M. Zalacain, A. Jiménez.** (1989). Molecular analysis

of the pac gene encoding a puromycin N-acetyl transferase from *Streptomyces alboniger*. Gene **79**, 375-380.

84. **Lam, W., and W. F. Doolittle.** (1989). Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA **86**, 5478-5482.
85. **Mankin, A. S., I. M. Zyrianova, V. K. Kagramanova, and R. A. Garrett.** (1992). Introducing mutations into the single-copy chromosomal 23S rRNA gene of the archaeon *Halobacterium halobium* by using an rRNA operon-based transformation system. Proc. Natl. Acad. Sci. USA **89**, 6535-6539.
86. **Martin, A., S. Yeats, D. Janekovic, W.-D. Reiter, W. Aicher, and W. Zillig.** (1984). SAV1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J. **3**, 2165-2168.
87. **Mazodier, P., and J. Davies.** (1991). Gene transfer between distantly related organisms. Annu. Rev. Genet. **25**, 147-171.
88. **Meile, L., A. Kiener, and T. Leisinger.** (1983). A plasmid in the archaebacterium *Methanobacterium thermoautotrophicum*. Mol. Gen. Genet. **191**, 480-484.
89. **Meile, L., T. Leisinger, and J. N. Reeve.** (1985). Cloning of DNA sequences from *Methanococcus vannielii* capable of autonomous replication in yeast. Arch. Microbiol. **143**, 253-255.
90. **Meile, L., and J. N. Reeve.** (1985). Potential shuttle vectors based on the methanogen plasmid pME2001. BioTechnology **3**, 69-72.
91. **Meile, L., J. Mazdon, and T. Leisinger.** (1988). Identification of a transcript and its promoter region on the archaebacterial plasmid pME2001. J. Bacteriol. **170**, 478-481.
92. **Meile, L., U. Jenal, D. Studer, M. Jordan, and T. Leisinger.** (1989). Characterization of ψM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. **152**, 105-110.
93. **Meile, L., P. Abendschein, and T. Leisinger.** (1990). Transduction in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. J. Bacteriol. **172**, 3507-3508.
94. **Mevarech, M., and R. Werczberger.** (1985). Genetic transfer in *Halobacterium volcanii*. J. Bacteriol. **162**, 461-462.
95. **Micheletti, P. A., K. A. Sment, and J. Konisky.** (1991). Isolation of a coenzyme M-auxotrophic mutant and transformation by electroporation in *Methanococcus voltae*. J. Bacteriol. **173**, 3414-3418.
96. **Montero, C. G., A. Ventosa, J. J. Nieto, and F. Ruiz-Berraquero.** (1988). Isolation and partial characterization of a plasmid in the extremely halophilic archaebacterium *Halococcus morrhuae* CCM 537. J. Gen. Microbiol. **134**, 2959-2963.
97. **Nadel, M., G. Mirambeau, P. Forterre, W.-D. Reiter, and M. Duguet.** (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature (London) **321**, 256-258.
98. **Neumann, H., V. Schwass, C. Eckerskorn, and W. Zillig.** (1990). Identification and characterization of the genes encoding the three structural proteins of the *Thermoproteus tenax* virus TTV1. Mol. Gen. Genet. **217**, 105-110.
99. **Neumann, H., and W. Zillig.** (1990). Structural variability in the genome of the *Thermoproteus tenax* virus TTV1. Mol. Gen. Genet. **222**, 435-437.
100. **Ng, W., S. Kothakota, and S. DasSarma.** (1991). Structure of the gas vesicle plasmid in *Halobacterium halobium*: inversion isomers, inverted repeats, and insertion sequences. J. Bacteriol. **173**, 1958-1964.
101. **Nieuwlandt, D. T., and C. J. Daniels.** (1990). An expression vector for the archaebacterium *Haloferax volcanii*. J. Bacteriol. **172**, 7104-711
102. **Nölling, J.** (1986). Untersuchungen zur Kopienzahl und Stabilität des Plasmids pME2001 in *Methanobacterium thermoautotrophicum* Stamm Marburg. Diplomarbeit, Philipps University of Marburg, Germany.

103. **Nölling, J., M. Frijlink, and W. M. de Vos.** (1991). Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. J. Gen. Microbiol. **137**, 1981-1986 (Chapter 4).
104. **Nölling J., and W. M. de Vos** (1992). Characterization of the archaeal, plasmid-encoded type II restriction-modification system *Mth*TI from *Methanobacterium thermoformicicum*: homology to the bacterial *Ngo*PII system from *Neisseria gonorrhoeae*. J. Bacteriol. **174**, 5719-5726 (Chapter 7).
105. **Nölling, J., and W. M. de Vos.** (1992). Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene. Nucleic Acids Res. **20**, 5047-5052 (Chapter 8).
106. **Nölling, J., F. J. M. van Eeden, R. I. L. Eggen, and W. M. de Vos.** (1992). Modular organization of related archaeal plasmids encoding different restriction-modification systems in *Methanobacterium thermoformicicum*. Nucleic Acids Res. **20**, 6501-6507 (Chapter 5).
107. **Nölling, J., F. J. M. van Eeden, and W. M. de Vos.** (1993). Distribution and characterization of plasmid-related sequences in the chromosomal DNA of different thermophilic *Methanobacterium* strains. Mol. Gen. Genet., accepted for publication (Chapter 6).
108. **Nölling, J., A. Groffen, and W. M. de Vos.** (1993). ΦF1 and ΦF3, two novel virulent archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. Submitted for publication (Chapter 3).
109. **Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W.-D. Reiter, and W. Zillig.** (1991). Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185**, 242-250.
110. **Pauling, C.** (1982). Bacteriophages of *Halobacterium halobium*: isolation from fermented fish sauce and primary characterization. Can. J. Microbiol. **28**, 2165-2168.
111. **Perler, F. B., D. G. Comb, W. E. Jack, L. S. Moran, B. Qiang, R. B. Kucera, J. Benner, B. E. Slatko, D. O. Nwankwo, S. K. Hempstead, C. K. S. Carlow, and H. Jannasch.** (1992). Intervening sequences in an archaea DNA polymerase gene. Proc. Natl. Acad. Sci. **89**, 5577-5581.
112. **Pfeifer, F., G. Weidinger, and W. Goebel.** (1981). Characterization of plasmids in halobacteria. J. Bacteriol. **145**, 369-374.
113. **Pfeifer, F., G. Weidinger, and W. Goebel.** (1981). Genetic variability in *Halobacterium halobium*. J. Bacteriol. **145**, 375-381.
114. **Pfeifer, F., M. Betlach, G. Weidinger, and W. Goebel.** (1982). Structures and functions of chromosomal and extrachromosomal DNA in halobacteria. Zbl. Bakt. Hyg., I. Abt. Orig. C **3**, 110-119.
115. **Pfeifer, F., M. Betlach, R. Martienssen, J. Friedman, and H. W. Boyer.** (1983). Transposable elements in *Halobacterium halobium*. Mol. Gen. Genet. **191**, 182-188.
116. **Pfeifer, F., J. Friedman, H. W. Boyer, and M. Betlach.** (1984). Characterization of insertions affecting the expression of the bacterio-opsin gene in *Halobacterium halobium*. Nucleic Acids Res. **12**, 2489-2497.
117. **Pfeifer, F., H. B. Boyer, and M. Betlach.** (1985). Restoration of bacterioopsin gene expression in a revertant of *Halobacterium halobium*. J. Bacteriol. **164**, 414-420.
118. **Pfeifer, F.** (1986). Insertion elements and genome organization of *Halobacterium halobium*. System. Appl. Microbiol. **7**, 36-40.
119. **Pfeifer, F., U. Blaseio, and P. Ghahraman.** (1988). Dynamic plasmid populations in *Halobacterium halobium*. J. Bacteriol. **170**, 3718-3724.
120. **Pfeifer, F., and U. Blaseio.** (1990). Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. Nucleic Acids Res. **18**, 6921-6925.
121. **Pfeifer, F., and P. Ghahraman.** (1991). The halobacterial insertion element ISH28. Nucleic Acids Res. **19**, 5788.

122. **Reeve, J. N.** (1992). Molecular biology of methanogens. Annu. Rev. Microbiol. **46**, 165-191.
123. **Reiter, W.-D., P. Palm, S. Yeats, and W. Zillig.** (1987). Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. **209**, 270-275.
124. **Reiter, W.-D., P. Palm, and S. Yeats.** (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17**, 1907-1914.
125. **Rosenshine, I., R. Tchelet, and M. Mevarech.** (1988). The mechanism of DNA transfer in the mating system of an archaebacterium. Science **245**, 1387-1389.
126. **Rosenshine, I., and M. Mevarech.** (1989). Isolation of plasmids found in three *Halobacterium volcanii* species. Can. J. Microbiol. **35**, 92-95.
127. **Ross, H. N. M., and W. D. Grant.** (1985). Nucleic acid studies on halophilic archaebacteria. J. Gen. Microbiol. **131**, 165-173.
128. **Sandbeck, K. A., and J. A. Leigh.** (1991). Recovery of an integration shuttle vector from tandem repeats in *Methanococcus maripaludis*. Appl. Environ. Microbiol. **52**, 2762-2763.
129. **Sandman, K., J. A. Krzycki, B. Dobrinsky, R. Lurz, and J. N. Reeve.** (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc. Natl. Acad. Sci. USA **87**, 5788-5791.
130. **Schleper, C., K. Kubo, and W. Zillig.** (1992). The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proc. Natl. Acad. Sci. USA **89**, 7645-7649.
131. **Schnabel, H., W. Zillig, M. Pfäffle, R. Schnabel, H. Michel, and H. Delius.** (1982). *Halobacterium halobium* phage ΦH. EMBO J. **1**, 87-92.
132. **Schnabel, H., E. Schramm, R. Schnabel, and W. Zillig.** (1982). Structural variability in the genome of phage ΦH of *Halobacterium halobium*. Mol. Gen. Genet. **188**, 370-377.
133. **Schnabel, H.** (1984). An immune strain of *Halobacterium halobium* carries the invertible L segment of phage ΦH as a plasmid. Proc. Natl. Acad. Sci. USA **81**, 1017-1020.
134. **Schnabel, H., P. Palm, K. Dick, and B. Grampp.** (1984). Sequence analysis of the insertion element ISH1.8 and of associated structural changes in the genome of phage ΦH of the archaebacterium *Halobacterium halobium*. EMBO J. **3**, 1717-1722.
135. **Schnabel, H., and W. Zillig.** (1984). Circular structure of the genome of phage ΦH in a lysogenic *Halobacterium halobium*. Mol. Gen. Genet. **193**, 422-426.
136. **Simon, R. D.** (1978). *Halobacterium* strain 5 contains a plasmid which is correlated with the presence of gas vacuoles. Nature (London) **273**, 314-317.
137. **Simon, R. D.** (1981). Morphology and protein composition of gas vesicles from the wild type and gas vacuole deficient strains of *Halobacterium salinarium* strain 5. J. Gen. Microbiol. **125**, 103-111.
138. **Simsek, M., S. DasSarma, U. L. RajBhandary, and H. G. Khorana.** (1982). A transposable element from *Halobacterium halobium* which inactivates the bacteriorhodopsin gene. Proc. Natl. Acad. Sci. USA **79**, 7268-7272.
139. **Sioud, M., G. Baldacci, P. Forterre, and A.-M. de Recondo.** (1988). Novobiocin induces accumulation of a single strand plasmid pGRB-1 in the archaebacterium *Halobacterium* GRB. Nucleic Acids Res. **16**, 7833-7842.
140. **Sowers, K. R., and R. P. Gunsalus.** (1988). Plasmid DNA from the acetotrophic methanogen *Methanosarcina acetivorans*. J. Bacteriol. **170**, 4979-4982.
141. **Sprague, G. F.** (1991). Genetic exchange between kingdoms. Current Opinion in Genetics and Development **1**, 530-533.
142. **Stax, D., R. Hermann, R. Falchetto, and T. Leisinger.** (1992). The lytic enzyme in bacteriophage ψM1-induced lysates of *Methanobacterium thermoautotrophicum* Marburg. FEMS Microbiol. Lett. **100**, 433-438.
143. **Stetter, K. O., and W. Zillig.** (1985). *Thermoplasma* and the thermophilic sulfur-dependent archaebacteria. In: The Bacteria., Vol. 8, Archaebacteria, p. 85-170. (eds. C. R. Woese and

R. S. Wolfe), Academic Press, Orlando, Fl.
144. **Thomm, M., J. Altenbuchner, and K. O. Stetter.** (1983). Evidence for a plasmid in a methanogenic bacterium. J. Bacteriol. **153**, 1060-1062.
145. **Tindall, B. J., H. N. M. Ross, and W. D. Grant.** (1984). *Natronobacterium* gen. nov. and *Natronococcus* gen. nov., two new genera of haloalkaliphilic archaebacteria. System. Appl. Microbiol. **5**, 41-57.
146. **Tindall, B. J.** (1992). The family *Halobacteriaceae*. In: The Prokaryotes, 2nd ed., p.768-808 (eds. A. Balows, H. G. Trüper, M. Dworkin, W. Harder and K.-H. Schleifer), Springer Verlag, New York.
147. **Torsvik, T., and I. D. Dundas.** (1974). Bacteriophage of *Halobacterium halobium*. Nature (London) **248**, 680-681.
148. **Torsvik, T., and I. D. Dundas.** (1980). Persisting phage infection in *Halobacterium salinarium* str. 1. J. Gen. Virol. **47**, 29-36.
149. **Torsvik, T.** (1982). Characterization of four bacteriophages for *Halobacterium* with emphasis on phage Hs1. In: Archaebacteria (eds. O. Kandler), Gustav Fischer Verlag, Stuttgart-New York.
150. **Vogels, G. D. , W. F. Hoppe, and C. K. Stumm.**(1980). Association of methanogenic bacteria with rumen ciliates. Appl. Environ. Microbiol. **40**, 608-612.
151. **Vogelsang-Wenke, H., and D. Oesterhelt.** (1988). Isolation of a halobacterial phage with a fully cytosine-methylated genome. Mol. Gen. Genet. **211**, 407-414.
152. **Wais, A. C., M. Kon, R. E. MacDonald, and B. D. Stollar.** (1975). Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H.halobium*. Nature (London) **256**, 314-315.
153. **Weidinger, G., G. Klotz, and W. Goebel.** (1979). A large plasmid from *Halobacterium halobium* carrying genetic information for gas vacuole formation. Plasmid **2**, 377-386.
154. **Whitman, W. B., T. L. Bowen, and D. R. Boone.** (1992). The methanogenic bacteria. In: The Prokaryotes, 2nd ed., p.719-767 (eds. A. Balows, H. G. Trüper, M. Dworkin, W. H. Harder, K. H. Schleifer), Springer Verlag, New York.
155. **Wilharm, T., M. Thomm, and K. O. Stetter.** (1986). Genetic analysis of plasmid pMP1 from *Methanolobus vulcani*. System. Appl. Microbiol. **7**, 401.
156. **Wittmann-Liebold, B., A. K. E. Köpke, E. Arndt, W. Krömer, T. Hatakeyama, H. G. Wittmann.** (1990). Sequence comparison and evolution of the ribosomal proteins and their genes. In: The Ribosome. Structure, Function and Evolution, p. 598-616 (ed.W. E. Hill), American Society for Microbiology, Washington D.C.
157. **Woese, C. R.** (1987). Bacterial evolution. Microbiol. Rev. **51**, 221-271.
158. **Woese, C. R., O. Kandler, and M. L. Wheelis.** (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87**, 4576-4579.
159. **Woese, C. R.** (1992). Prokaryote systematics: the evolution of a science. In: The Prokaryotes, 2nd ed., p. 3-18 (eds. A. Balows, H. G. Trüper, M. Dworkin, W. H. Harder, K. H. Schleifer), Springer Verlag, New York.
160. **Wood, A. G., W. B. Whitman, and J. Konisky.** (1985). A newly-isolated marine methanogen harbors a small cryptic plasmid. Arch. Microbiol. **142**, 259-261.
161. **Wood, A. G., W. B. Whitman, and J. Konisky.** (1989). Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. J. Bacteriol. **171**, 93-98.
162. **Worrel, V. E., D. P. Nagle, D. McCarthy, and A. Eisenbran.** (1988). Genetic transformation system in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. J. Bacteriol. **170**, 653-656.
163. **Xu, W. L., and W. F. Doolittle.** (1983). Structure of the archaebacterial insertion element ISH50. Nucleic Acids Res. **11**, 4195-4199.
164. **Yeats, S., P. McWilliam, and W. Zillig.** (1982). A plasmid in the archaebacterium

Sulfolobus acidocaldarius. EMBO J. **1**, 1035-1038.

165. **Zabel, H. P., E. Holler, and J. Winter.** (1987). Mode of inhibition of the DNA polymerase of *Methanococcus vannielii* by aphidicolin. Eur. J. Biochem. **165**, 171-175.
166. **Zhao, H., A. G. Wood, F. Widdel, and M. Bryant.** (1988). An extremely thermophilic *Methanococcus* from a deep sea hydrothermal vent and its plasmid. Arch. Microbiol. **150**, 178-183.
167. **Zillig, W., S. Yeats, I. Holz, A. Böck, F. Gropp, M. Rettenberger, and S. Lutz.** (1985). Plasmid-related anaerobic autotrophy of the novel archaebacterium *Sulfolobus ambivalens*. Nature (London) **313**, 789-791.
168. **Zillig, W., S. Yeats, I. Holz, A. Böck, M. Rettenberger, F. Gropp, and G. Simon.** (1986) *Desulfurolobus ambivalens*, gen. nov., sp. nov., an autotrophic archaebacterium facultatively oxidizing or reducing sulfur. System. Appl. Microbiol. **8**, 197-203.
169. **Zillig, W., F. Gropp, A. Henschen, H. Neumann, P. Palm, W.-D. Reiter, M. Rettenberger, H. Schnabel, and S. Yeats.** (1986). Archaebacterial virus host systems. System. Appl. Microbiol. **7**, 58-66.
170. **Zillig, W., I. Holz, H.-P. Klenk, J. Trent, S. Wunderl, D. Janeković, E. Imsel, and B. Haas.** (1987). *Pyrococcus woesei* sp. nov., an ultra-thermophilic marine archaebacterium, representing a novel order, *Thermococcales*. System. Appl. Microbiol. **9**, 62-70.
171. **Zillig, W., P. Palm, W. D. Reiter, G. Pühler, and H.-P. Klenk.** (1988). Comparative evaluation of gene expression in archaebacteria. Eur. J. Biochem. **173**, 473-482.
172. **Zillig, W., W.-D. Reiter, P. Palm, F. Gropp, H. Neumann, and M. Rettenberger.** (1988). Viruses of archaebacteria. In: The Viruses, Vol 1, p. 517-558 (ed. R. Calendar), Plenum Press, New York.
173. **Zillig, W.** (1991). Comparative biochemistry of *Archaea* and *Bacteria*. Current Opinion in Genetics and Development **1**, 544-551.
174. **Zillig, W., P. Palm, H.-P. Klenk, G. Pühler, F. Gropp, and C. Schleper.** (1991). Phylogeny of DNA-dependent RNA polymerases: Testimony for the origin of eukaryotes. In: General and Applied Aspects of Halophilic Microorganisms, p. 321-332 (ed. F. Rodriguez-Valera), Plenum Press, New York/London.
175. **Zvyaga, T. A., S. A. Zozulya, and S. O. Guriev.** (1987). The nucleotide sequence of the archaebacterial transposable element ISH S1. Bioorg. Chem. **13**, 1351-1357.

Phylogenetic analysis of thermophilic *Methanobacterium* sp.: evidence for a common formate-utilizing ancestor

Jörk Nölling, Dittmar Hahn, Wolfgang Ludwig[1] and Willem M. de Vos

[1]Lehrstuhl für Mikrobiologie, Technische Universität München, Federal Republic of Germany.

SUMMARY

Comparative sequence analysis of PCR-amplified 16S rRNA genes from members of *M.thermoformicicum* and *M.thermoautotrophicum* provide evidence that the classification into these two species does not reflect their phylogenetic relationship. Our results support the proposal of Touzel and coworkers (1992) who, based on DNA hybridization and immunological studies, identified at least three groups among strains of *M.thermoformicicum* and *M.thermoautotrophicum* which may represent different species. Those three groups comprise (i) *M.thermoformicicum* strains Z-245, FTF, THF, CSM3, FF1 and FF3, and *M.thermoautotrophicum* ΔH, (ii) *M.thermoformicicum* strains CB12, SF-4 and HN4, and (iii) *M.thermoautotrophicum* Marburg. In addition, sequences related to the *fdh*A and *fdh*B genes for formate dehydrogenase were identified in the genomes of *M.thermoautotrophicum* strains ΔH and Marburg suggesting that *M.thermoautotrophicum* and *M.thermoformicicum* have been derived from a common ancestor able to utilize formate.

INTRODUCTION

The thermophilic species *Methanobacterium thermoautotrophicum* and *Methanobacterium thermoformicicum* are classified according to their ability to use H_2/CO_2 only or H_2/CO_2 and formate, respectively, as sole energy- and carbon source (Whitman et al., 1992). Several observations suggest, however, that a taxonomy based on a phenotypical character such as formate-utilization may not reflect the phylogenetic relationship of the two species. The assignment of *M.thermoautotrophicum* strains ΔH and Marburg into the same species was questioned already more than ten years ago by Brandis et al. (1981) who found that the strains ΔH and Marburg exhibit only limited (46%) DNA-DNA relatedness and show considerable differences in composition of cell wall amino sugars and the large subunit of the DNA dependent RNA polymerase. *M.thermoformicicum* strains, on the other hand, differ in respect to immunological cross-reactions (Yamamoto et al., 1989) and genetic fingerprinting (Nölling et al., 1991). Moreover, comparison of 5S rRNA sequences (Chumakov et al., 1987) and immunological studies (Bezrukova et al., 1989) indicate that *M.thermoautotrophicum* ΔH is closer related to *M.thermoformicicum* Z-245 than to *M.thermoautotrophicum* Marburg.

The question whether the strains of *M.thermoautotrophicum* and *M.thermoformicicum* have to be reclassified should be answered by a detailed analysis of their similarity on nucleic acid level, e.g. by DNA-DNA hybridization studies and ribosomal RNA sequence comparisons. The recently reported results from DNA-DNA similarity values and immunological studies of several strains of the species *M.thermoautotrophicum* and *M.thermoformicicum* (Touzel et al., 1992) clearly indicate that reclassification of both species is required. It was found that (i) *M.thermoautotrophicum* ΔH is only distantly related to *M.thermoautotrophicum* Marburg but more related to some members of the species *M.thermoformicicum*, and (ii) the species *M.thermoformicicum* includes strains which may belong to another, not yet defined species (Touzel et al., 1992).

One of the powerful methods to determine the evolutionary relationship among different organisms is the analysis of the variable regions of small subunit rRNAs sequences (Woese et al., 1987). Among the thermophilic strains of the genus *Methanobacterium*, only the rRNA operon of *M.thermoautotrophicum* Marburg has so far been characterized in detail by sequence analysis of a genomic clone (Østergaard et al., 1987). The aim of the present report was to analyze the differences in the 16S rRNA of members of *Methanobacterium* in order to confirm their inter- and intraspecies relationship derived from DNA-DNA similarity studies (Touzel et al., 1992). For this purpose, we determined the partial 16S rRNA sequences of *M.thermoautotrophicum* ΔH and *M.thermoformicicum* strains Z-245, FTF, THF, CB12, SF-4, FF1, FF3, CSM3 and HN4 from cloned 16S rRNA genes that were amplified by the polymerase chain reaction (PCR), and compared these sequences with that of the 16S rRNA from *M.thermoautotrophicum* Marburg. Furthermore, we show that the non-formate utilizing strains *M.thermoautotrophicum* Marburg and ΔH contain sequences related to the genes encoding the α and β subunit of the formate-dehydrogenase (*fdh*A and *fdh*B, respectively) from *M.formicicum*.

MATERIAL AND METHODS

Methanogenic strains

The following strains were obtained from the German Collection of Microorganisms (Braunschweig, Germany): *M.thermoautotrophicum* ΔH (DSM 1053), *M.thermoformicicum* strains Z-245 (DSM 3720), FTF (DSM 3012), THF (DSM 3848) and CB12 (DSM 3664). *M.thermoformicicum* SF-4, kindly provided by K. Yamamoto, Japan, and strains FF1 and FF3 have been described previously (Nölling et al., 1991). *M.thermoformicicum* strains CSM3 and HN4 were isolated according to previously described procedures (Nölling et al., 1991) from a mesophilic upflow-anaerobic-sludge-bed reactor (T. Grotenhuis and C. Plugge, Department of Microbiology, University of Wageningen, The Netherlands) and from sediment of the river Rhine collected nearby Wageningen, The Netherlands, respectively.

PCR amplification of 16S rRNA-genes

Total DNA from methanogenic bacteria was isolated as described previously (Nölling et al., 1991).

To facilitate cloning of PCR products, two primers commonly used for reverse transcriptase sequencing (Embley et al., 1988) were synthesized with a 5' extension containing an unique

restriction enzyme cleavage sites. Primer 1510 (Embley et al., 1988) was equipped with a *Pst*I-site resulting in primer 1510Pst with the sequence 5'-GTG*CTGCAG*GGTTACCTTGTTACGACT (complementary to position 1493 to 1510 in the *Escherichia coli* numbering system). The second primer, 124Bam, complementary to primer 124 (Embley et al., 1988), contained an additional *Bam*HI-site and had the sequence 5'-CAC*GGATCC*GGACGGGTGAGTAACACG (position 106 to 124 in the *E.coli* numbering system).

PCR was performed in a total volume of 100 µl containing 10 µl 10 x PCR buffer (500 mM KCl, 25 mM $MgCl_2$, 200 mM Tris/HCl pH 8.4, 1 mg/ml gelatine, 0.001% Nonidet P-40), 1 µl dNTPs (each 10 mM in 10mM Tris/HCl pH 7.5), 0.2 µl Taq polymerase (5 U/µl, Life Technologies Inc, Gaithersburg, Md.), 1 µl of each of the primers 1510Pst and 124Bam (100 ng/µl), and 1 µl DNA preparation (approximately 100 ng). Thirty rounds of temperature cycling (95°C for 1 min, 48°C for 2 minutes and 72°C for 3 minutes) were followed by a final 7 min step at 72°C.

Cloning and sequencing of amplified PCR products

The amplification products were treated with proteinase K (0.1 mg/ml, Boehringer Mannheim, FRG) according to Crowe et al. (1991) except that sodium dodecyl sulfate (SDS) was added to a final concentration of 0.5% followed by an incubation at 55°C for 20 min. Subsequently, the suspension was extracted with phenol/chloroform, and the DNA was precipitated with ethanol and finally dissolved in TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). The amplified DNA was digested with *Bam*HI and *Pst*I (Life Technologies Inc., Gaithersburg, Md.), separated by agarose gel electrophoresis and purified by the Geneclean procedure (Bio 101, La Jolla, Calif.). Cloning of the amplification products into *Bam*HI/*Pst*I-digested phage M13mp18/19 (Yanisch-Perron et al., 1985) and isolation of double- and single-stranded M13-DNA was done by standard methods (Sambrook et al., 1989), using *E.coli* TG1 (Gibson, 1983) as host.

The nucleotide sequences of overlapping fragments derived from the cloned amplification products were determined by the dideoxy chain termination method (Sanger et al., 1977) using the Sequenase kit according to the manufacturer's instructions (United States Biochemical Corp. Cleveland, Ohio). In addition to the supplied M13-forward sequencing primer, two new primers were synthesized for determination of the internal sequence of the cloned amplification products: primer 361 with the sequence 5'-GGGGCGCAGCAGGCGC corresponding to positions 346 to 361 (*E.coli* numbering system) and primer 1114 with the sequence 5'-GGGTCTCGCTCGTTGCC complementary to positions 1098 to 1114 (*E.coli* numbering system). The 16S rRNA sequence of each strain used for comparison was based on two different clones derived from the same PCR reaction, or - in the case of *M.thermoformicicum* strains Z-245, THF, CSM3, FF1 and HN4, and *M.thermoautotrophicum* ΔH - from independent PCR reactions.

DNA hybridization

For the detection of formate-dehydrogenase-related sequences present in the genomes of different *Methanobacterium* strains a radioactively labelled 2.2-kb *Bam*HI-*Sph*I fragment containing the *fdh*A gene or a 1.8-kb *Bam*HI-*Sph*I fragment containing the *fdh*B gene from *M.formicicum* JF-1 (Shuber et al., 1986; kindly provided by J. G. Ferry, Department of

Anaerobic Microbiology, VPI Blacksburg, VA 24061, USA) were used as probes. Total DNA prepared from *Methanobacterium* strains was digested, electrophoresed on a 0.7% agarose gel, and transferred to a Hybond-N nylon filter (Amersham, United Kingdom) by the capillary blot procedure (Sambrook et al., 1989). Prehybridization of the nylon membranes was performed at 65°C for 15 min in 0.5 M $NaHPO_4$ pH 7.2, 1 mM EDTA, 7% SDS, 1% serum albumin and 0.5 mg sonicated and denaturated salmon sperm DNA. After purification from agarose gel, the DNA probe was radioactively labelled by nick-translation (Sambrook et al., 1989) with [α-^{32}P]dATP (110 TBq/mmol; Amersham, United Kingdom), denaturated and added to the prehybridization solution. Hybridization was performed at 55°C for 18 h. The membranes were washed for 5 min at room temperature in 2 x SSC (0.3 M NaCl, 0.03 M sodium citrate) containing 0.1% SDS, followed by incubation in the same buffer for 20 min at 55°C and were finally subjected to autoradiography.

RESULTS AND DISCUSSION

Hybridization with *fdh* genes

Among the thermophilic species of the order *Methanobacteriales*, only *M.thermoformicicum* is able to convert formate, whereas *M.thermoautotrophicum* strains have not been reported to utilize formate (Whitman et al., 1992). We have previously shown that comparison of hybridization patterns derived from probing with *fdh*A, the structural gene coding for the α subunit of formate hydrogenase of the mesophilic *M.formicicum* (Shuber et al., 1986), is a useful technique for the classification of *M.thermoformicicum*. Two groups could be distinguished using this approach: the Z-245-group comprising strains Z-245, FTF, THF, FF1, and FF3, and the CB12-group including strains CB12 and SF-4 (Nölling et al., 1991). The two new *M.thermoformicicum* isolates CSM3 and HN4 were classified in the same way (data not shown) and found to belong to the Z-245-group and the CB12-group, respectively.

Unexpectedly, we found that radioactively labelled *fdh*A hybridized also with total DNA of the non-formate utilizers *M.thermoautotrophicum* strains ΔH and Marburg (Fig. 1A). Moreover, *M.thermoautotrophicum* strains ΔH and Marburg also contained genomic sequences homologous to *fdh*B (Fig. 1B), which encodes the β subunit of the *M.formicicum* formate dehydrogenase (Shuber et al., 1986). The hybridization signals obtained with radioactively labelled *fdh*A and *fdh*B revealed a hierarchy of intensity in which the most intense signals were found with DNA from *M.thermoformicicum*, while the signals of *M.thermoautotrophicum* ΔH DNA were stronger than those with DNA from *M.thermoautotrophicum* Marburg (Fig. 1).

The hybridization data clearly show that the genomes of *M.thermoautotrophicum* strains ΔH and Marburg contain *fdh*-related sequences suggesting that the inability of both strains to convert formate is due to a non-functionality of the *fdh*-operon rather than the absence of the structural genes for formate dehydrogenase. If so, the non-functionality of the *fdh* genes in the *M.thermoautotrophicum* strains may explain the observed less intense hybridization signals compared to those obtained with DNA from members of the only distantly related formate-utilizing *M.thermoformicicum* Z-245- and CB12-group (Fig. 1): the selective pressure on the latter strains should ensure conservation of the *fdh* genes whereas mutations in the non-essential *fdh*-related sequences of *M.thermoautotrophicum* strains ΔH and Marburg would not have been selected out.

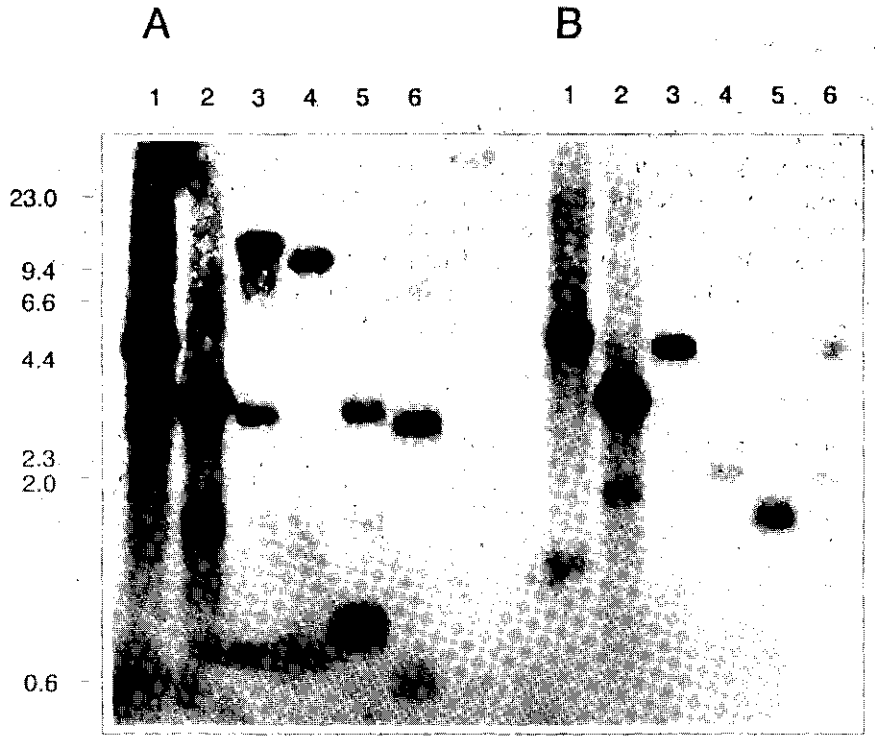


Figure 1. Southern hybridization of total DNA isolated from *Methanobacterium* with the *M.formicicum fdh*A (panel A) and *fdh*B genes (panel B). Total DNA (about 0.5 μg each) was isolated from *M.thermoformicicum* FTF (lane 1) and SF-4 (lane 2), and *M.thermoautotrophicum* strains ΔH (lane 3, 5) and Marburg (lane 4, 6), digested with *Bam*HI (lanes 1-4) and *Bam*HI/*Sph*I (lanes 5, 6) and separated by agarose gel electrophoresis. After transfer of the DNA to nylon-membranes, the blots were hybridized with a radioactively labelled 2.2-kb *Bam*HI-*Sph*I fragment containing the *fdh*A (panel A) and a 1.8-kb *Bam*HI-*Sph*I fragment containing the *fdh*B gene (panel B) from *M.formicicum* JF-1. The sizes (in kilobases) of *Hind*III λ DNA fragments are indicated at the left.

The presence of cryptic *fdh* genes implies that the strains ΔH and Marburg were derived from ancestors that were able to convert formate. *M.thermoautotrophicum* ΔH is closely related on genomic level to the *M.thermoformicicum* Z-245-group (Touzel et al., 1992; see below) and therefore represents a non-formate utilizing member of *M.thermoformicicum*. Strain Marburg, in contrast, is only distantly related to the thermophilic formate-utilizing *Methanobacterium* strains (Touzel et al., 1992; see below). Since both *M.thermoautotrophicum* strains harbor cryptic *fdh* genes, it is reasonable to assume that *M.thermoformicicum* and *M.thermoautotrophicum* strains ΔH and Marburg have been originated from a common formate-utilizing ancestor. Interestingly, a similar diversity in the ability to convert formate in addition to H_2/CO_2 is typical among mesophilic *Methanobacterium* strains and members of the closely related genus *Methanobrevibacter* (Whitman et al., 1992). It is therefore tempting to speculate that the non-formate utilizers among those strains also contain cryptic *fdh* genes. If so, this would imply that the capacity to convert formate into methane has occurred early in evolution of the genus *Methanobacterium* and its close relatives.

16S rRNA sequence analysis

Using PCR-amplified templates, partial 16S rRNA nucleotide sequences were determined from *M.thermoautotrophicum* ΔH and nine *M.thermoformicicum* strains including all seven strains reported so far and the two new isolates described here. The sequencing strategy generated a continuous stretch of 1334 bases of each strain. The 16S rRNA sequences from the nine strains examined were aligned with that of *M.thermoautotrophicum* Marburg (Ostergaard et al., 1987; Fig. 2). For calculation of the similarity values the 16S rRNA sequence of the mesophilic formate-utilizer M.formicicum was included (Table 1). Based on the derived K_{nuc} (evolutionary distance) values an unrooted phylogenetic tree was constructed

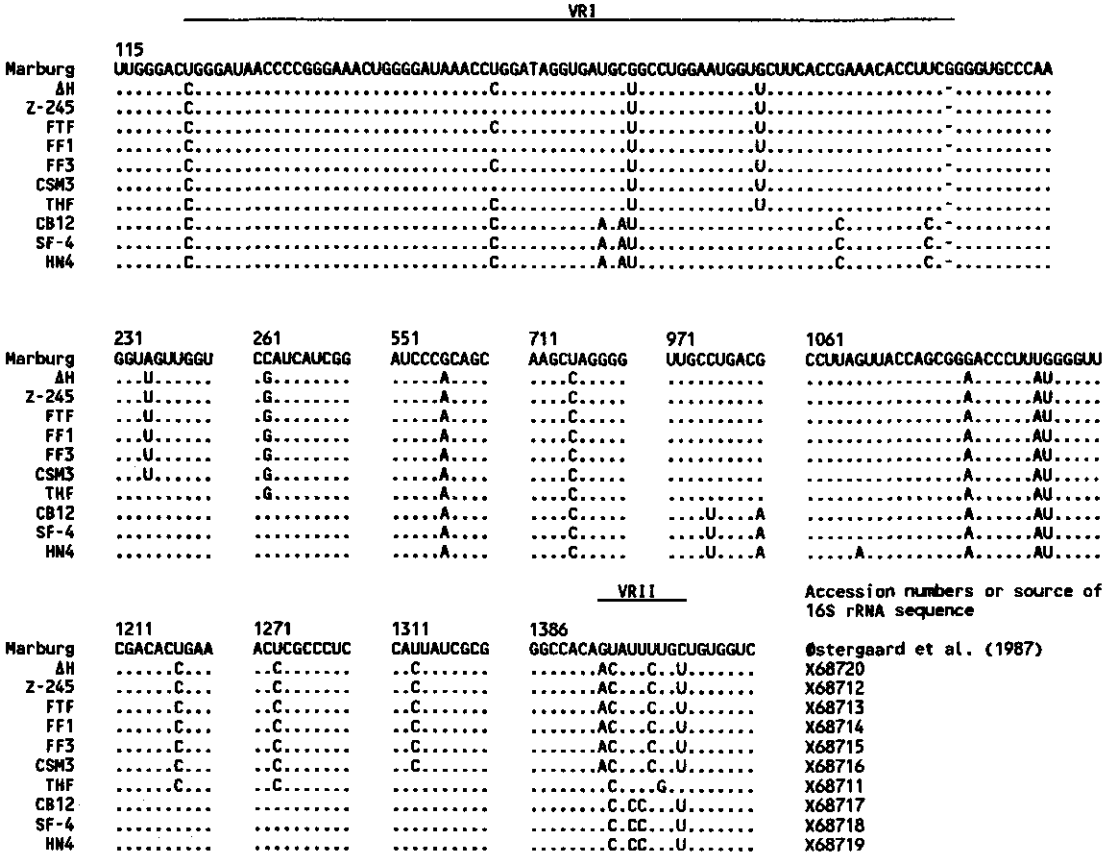


Figure 2. Alignment of variable regions of 16S rRNA sequences derived from *M.thermoautotrophicum* Marburg (Østergaard et al., 1987) and ΔH, and *M.thermoformicicum* strains Z-245, FTF, FF1, FF3, CSM3, THF, CB12, SF-4 and HN4. Base positions in the sequences are numbered consecutively and correspond to the numbering of the 16S rRNA from strain Marburg (Østergaard et al., 1987). Only nucleotides that differ from strain Marburg are shown, while identical nucleotides are denoted by dots. (-) correspond to alignment gaps. Regions VRI and VRII represent highly variable regions. The accession numbers (EMBL database) assigned to the analyzed 16S rRNA sequences are indicated.

using the neighbor-joining method of Saitou and Nei (1987) (Fig. 3).

The determined 16S rRNA sequences of the examined thermophilic *Methanobacterium* strains were highly similar ranging from 98.5 to 100%. In total 29 positions were found to be variable and a maximum of 19 base variations were observed between two sequences (Fig. 2). The majority (16) of those base variations was concentrated in two highly variable regions, designated VRI (position 122 to 200 according to the numbering of the 16S rRNA from *M.thermoautotrophicum* Marburg [Østergaard et al., 1987]) and VRII (position 1393 to 1401), which are located close to the 5'- and 3'-terminus of the 16S rRNA molecule, respectively (Figures 2 and 4). Alignment of the variable positions of the 16S rRNA molecules allowed the recognition of different 16S rRNA signatures and a classification of the corresponding *Methanobacterium* strains into at least three groups with nearly identical 16S rRNA sequences. One of these groups is represented by *M.thermoautotrophicum* Marburg. The two other groups are coinciding with those derived from genetic fingerprinting with *fdh*A (see above), i.e. the Z-245-group, comprising *M.thermoformicicum* strains Z-245, FTF, FF1, FF3 and CSM3, and the CB12-group, consisting of *M.thermoformicicum* strains CB12, SF-4 and HN4. Based on the 16S rRNA similarity, *M.thermoautotrophicum* ΔH belongs to the Z-245-group.

The 16S rRNA sequences of the Z-245- and the CB12-group differ from each other at 17 positions, corresponding to 98.7% sequence similarity. Both differ from *M.thermoautotrophicum* Marburg at 19 positions, corresponding to 98.6% sequence similarity. *M.thermoformicicum* THF has a slightly different 16S rRNA signature which shows similarity with those of all three groups. Its close proximity to the Z-245 branch suggests, however, that strain THF should be assigned to the Z-245 group as proposed on the basis of DNA-DNA hybridization studies (Touzel et al., 1992).

Using the secondary structural model of the 16S rRNA molecule of *M.thermoautotrophicum* Marburg (Østergard et al., 1987), the variable positions of the examined 16S rRNA sequences were analyzed for their possible effect on secondary structure. As illustrated in Figure 4 for the two highly variable regions VRI and VRII, about one third of the base

ble 1. Percentages of similarity (upper right) and K_{nuc} values (lower left) for a 1334 bp region of 16S rRNAs from strains of the genus *Methanobacterium*

Organism	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
. *M.thermoautotrophicum* Marburg	—	98.6	98.7	98.6	98.7	98.6	98.7	98.9	98.6	98.6	98.5	92.3
. *M.thermoautotrophicum* ΔH	0.0141	—	99.9	100.0	99.9	100.0	99.9	99.5	98.8	98.8	98.7	92.0
. *M.thermoformicicum* Z-245	0.0133	0.0001	—	99.9	100.0	99.9	100.0	99.5	98.7	98.7	98.6	92.0
. *M.thermoformicicum* FTF	0.0141	0.0	0.0001	—	99.9	100.0	99.9	99.5	98.8	98.8	98.7	92.1
. *M.thermoformicicum* FF1	0.0133	0.0001	0.0	0.0001	—	99.9	100.0	99.5	98.7	98.7	98.6	92.0
. *M.thermoformicicum* FF3	0.0141	0.0	0.0001	0.0	0.0001	—	99.9	99.5	98.8	98.8	98.7	92.1
. *M.thermoformicicum* CSM3	0.0133	0.0001	0.0	0.0001	0.0	0.0001	—	99.5	98.7	98.7	98.6	92.0
. *M.thermoformicicum* THF	0.0110	0.0045	0.0053	0.0045	0.0053	0.0045	0.0053	—	98.9	98.9	98.9	92.3
. *M.thermoformicicum* CB12	0.0141	0.0122	0.0129	0.0122	0.0129	0.0122	0.0129	0.0107	—	100.0	99.9	92.4
. *M.thermoformicicum* SF-4	0.0141	0.0122	0.0129	0.0122	0.0129	0.0122	0.0129	0.0107	0.0	—	99.9	92.4
. *M.thermoformicicum* HN4	0.0148	0.0129	0.0137	0.0129	0.0137	0.0129	0.0137	0.0114	0.0001	0.0001	—	92.3
. *M.formicicum* (DSM 1312)	0.0816	0.0841	0.0849	0.0832	0.0849	0.0831	0.0848	0.0808	0.0801	0.0801	0.0808	—

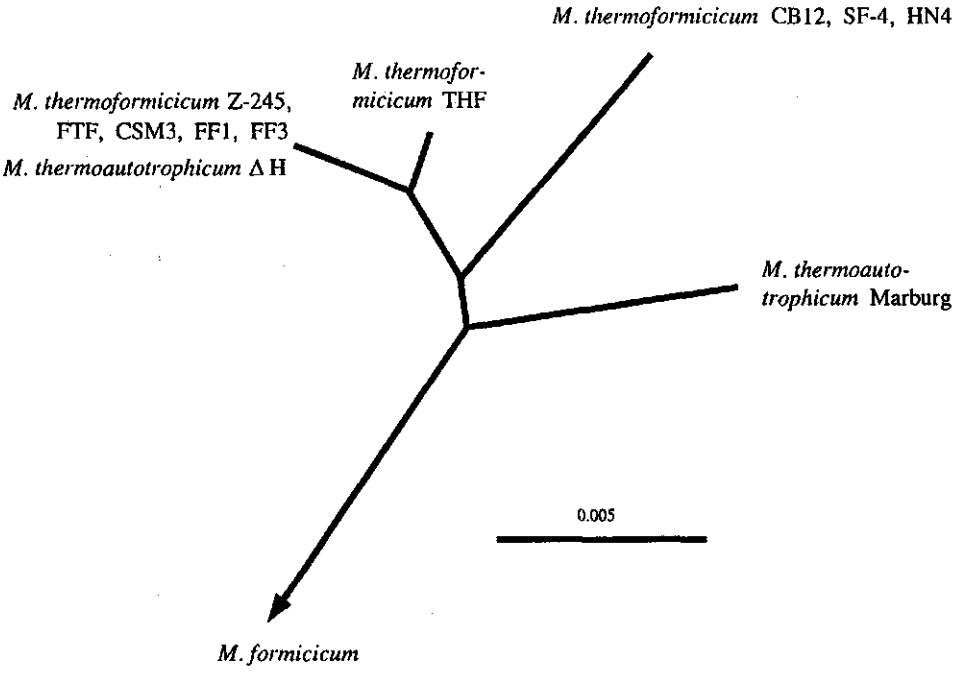


Figure 3. Phylogenetic tree reflecting the intrageneric relationships of *Methanobacterium* strains.

variations affected unpaired sequence regions that do not contribute to the stability of the secondary structure. In contrast, the other two third of the base substitutions were located in helical regions and most of those may affect the secondary structure of the 16S rRNA since only one position of a base pair was changed. Interestingly, the majority of those non-compensatory base substitutions affected mispaired regions in the 16S rRNA of *M.thermoautotrophicum* Marburg and generated perfect base pairing in the 16S rRNA sequences of members of the Z-245 and the CB12 group. Inspection of all variable positions located in helical regions of the analyzed 16S rRNAs of *M.thermoformicicum* and *M.thermoautotrophicum* ΔH revealed that this is true for the majority of base substitutions (Table 2).

Two conclusions may be drawn from the 16S rRNA sequence comparison. Firstly, the species *M.thermoformicicum* consists of two groups of strains, the Z-245- and the CB12-group, whose degree of phylogenetic relationship is as high as the relationship between each of these groups and *M.thermoautotrophicum* Marburg. Secondly, *M.thermoautotrophicum* ΔH is phylogenetically closer related to members of the Z-245-group than to *M.thermoautotrophicum* Marburg.

The obtained 16S rRNA data are in congruence with the results derived from DNA-DNA hybridization studies and antigenic fingerprinting described by Touzel et al. (1992) and support their conclusions that the examined strains of *M.thermoautotrophicum* and *M.thermoformicicum* comprise at least three groups of considerable similar strains. As shown in Table 3, the *M.thermoformicicum* strains Z-245 and CB12 on the one and *M.thermoautotrophicum* Marburg on the other hand exhibit about 30% DNA-DNA similarity. Based

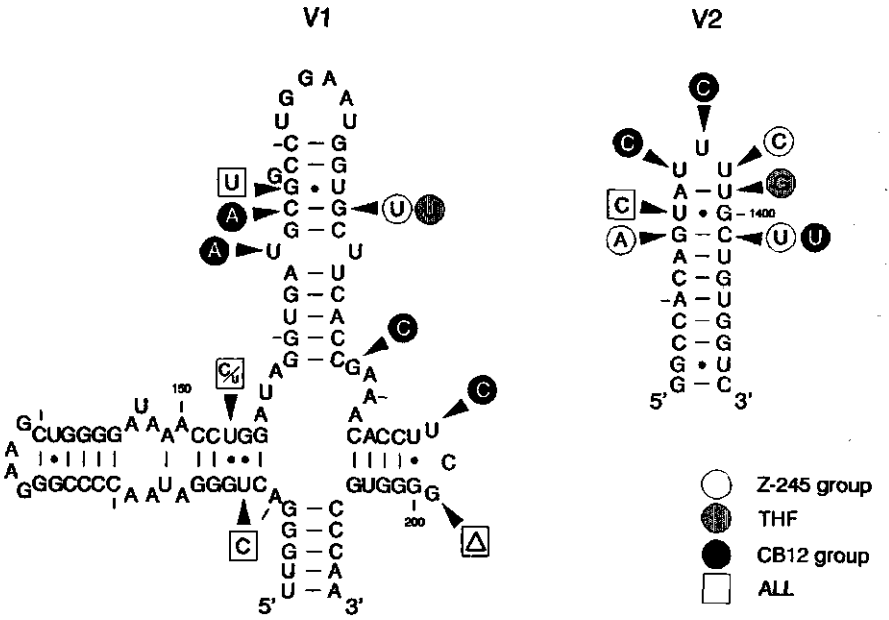


Figure 4. Secondary structure of the two variable regions VRI and VRII (see Fig. 2) of 16S rRNA sequences from strains of *M.thermoautotrophicum* and *M.thermoformicicum*. The secondary structure shown is that from *M.thermoautotrophicum* Marburg (Østergaard et al., 1987). Variable positions found in the 16S rRNA sequences from members of the Z-245- and CB12-group and from strain THF are indicated. A deleted nucleotide is denoted by (Δ). At position 153, the substitution (C) for (U) found for the majority of the strains is indicated.

Table 2. Variable base pairs in 16S rRNA from strains of the genus *Methanobacterium*

Position of base pair[a]	*M.thermoautotrophicum* Marburg	*M.thermoformicicum*		*M.formicicum*
		CB12-group	Z-245-group[b]	
122:154	U·G	C:G	C:G	U·G
123:153	G·U	G:C	G:Y[c]	G:C
164:182	U U	A:U	U U	G:C
166:180	C:G	A G	C U	C U
167:179	G·U	U U	U U	U U
556:568	G·U	A:U	A:U	A:U
715:752	U·G	C:G	C:G	C:G
955:975	G:C	G·U	G:C	A:U
1077:1099	G·U	A:U	A:U	G:C
1217:1222	U·G	U·G	C:G	U·G
1393:1401	G:C	G:C	A:U	G·U
1394:1400	U·G	C:G	C:G	U·G
1395:1399	A:U	A:U	A:G[d]	C:G

[a] Numbering follows *M.thermoautotrophicum* Marburg 16S rRNA.
[b] Including *M.thermoautotrophicum* ΔH.
[c] Y = pyrimidine.
[d] *M.thermoformicicum* THF only.

on these low values it was proposed that those strains belong to different species (Touzel et al. 1992). In contrast, the strains whose 16S rRNA sequences were found to be most similar to that of *M.thermoformicicum* Z-245, i.e. those of the Z-245-group, exhibit DNA-DNA similarity values of more than 70% (Table 3). According to Wayne et al. (1987), this is sufficiently high to assign these strains to the same species as proposed by Touzel and coworkers (1992). Although no DNA-DNA hybridization data are available that compare members of the CB12-group with each other, their nearly identical 16S rRNA signatures suggest that those strains belong to the same species.

Table 3. Comparison between DNA-DNA hybridization results (Touzel et al., 1992) and the levels of 16S rRNA sequence similarity

Organism pair	% of DNA-DNA hybridization	% of 16S rRNA homology
M.thermoformicicum Z-245 and *M.thermoformicicum* CB12	25	98.7
M.thermoformicicum CB12 and *M.thermoautotrophicum* Marburg	27	98.6
M.thermoformicicum CB12 and *M.thermoformicicum* THF	27	98.9
M.thermoformicicum Z-245 and *M.thermoautotrophicum* Marburg	29	98.6
M.thermoformicicum THF and *M.thermoautotrophicum* Marburg	33	98.9
M.thermoformicicum Z-245 and *M.thermoformicicum* THF	73	99.5
M.thermoformicicum Z-245 and *M.thermoautotrophicum* ΔH	97	99.9

The present 16S rRNA sequence analysis confirms the recently proposed classification of several thermophilic *Methanobacterium* strains based on DNA-DNA similarity studies (Touzel et al., 1992) and allows to determine the phylogenetic relationship of the examined methanogens. Additional support for the new classification is provided by the fact that the obtained genomic relatedness of the three thermophilic *Methanobacterium* groups correlates well with the established antigenic relationships (Touzel et al., 1992). Moreover, the homogeneity of the Z-245-group including *M.thermoautotrophicum* ΔH is supported by recent phage-typing studies using the newly identified *Methanobacterium* phage ΦF1 (Nölling et al., 1993a) and the presence of a specific DNA sequence possibly representing a mobile DNA element that was found to be restricted to strains of this group (Nölling et al., 1993b).

REFERENCES

Benstead, J., Archer, D. B., Lloyd, D.: Formate utilization by members of the genus *Methanobacterium*. Arch. Microbiol. *156*, 34-37 (1991)

Bezrukova, L. V., Obraztsova, A. Y., Zhilina, T. N.: Immunological studies on a group of methane-producing bacterial strains. Microbiology (Engl. Transl. Mikrobiologiya) *58*, 77-83 (1989)

Brandis, A., Thauer, R. K., Stetter, K. O.: Relatedness of strain ΔH and Marburg of *Methanobacterium thermoautotrophicum*. Zbl. Bakt. Hyg., I.Abt. Orig. C *2*, 311-317 (1981)

Crowe , J. S., Cooper, H. J., Smith, M. A., Sims, M. J., Parker, D., Gewert, D.: Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acids Res. *19*, 184 (1991)

Chumakov, K. M., Zhilina, T. N., Zvyagintseva, I. S., Tarasov, A. T., Zavarzin, G. A.: 5S RNA in archaebacteria. Zh. Obshch. Biol. *48*, 167-191 (in Russian) (1987)

Embley, T. M., Smida, J., Stackebrandt, E.: Reverse transcriptase sequencing of 16S ribosomal RNA from *Faenia rectivirgula*, *Pseudenocardia thermophilia* and *Saccharopolyspora hirsuta*, three wall type IV actinomyctes which lack mycolic acids. J. Gen. Microbiol. *134*, 961-966 (1988)

Gibson, T. J.: Ph.D thesis. University of Cambridge, United Kingdom (1983)

Lechner, K., Wich, G., Böck, A.: The nucleotide sequence of the 16S rRNA gene and flanking regions from *Methanobacterium formicicum*: the phylogenetic relationship between methanogenic and halophilic archaebacteria. System. Appl. Microbiol. *6*, 157-163 (1985)

Nölling, J., Frijlink, M., de Vos, W. M.: Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. J. Gen. Microbiol. *137*, 1981-1986 (1991) (Chapter 4)

Nölling, J., Groffen, A., de Vos, W. M.: ΦF1 and ΦF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. Submitted for publication (1993a) (Chapter 3)

Nölling, J., van Eeden, F. J. M., de Vos, W. M.: Distribution and characterization of plasmid-related sequences in the thermophilic DNA of different thermophilic *Methanobacterium* strains. Mol. Gen. Genet., accepted for publication (1993b) (Chapter 6)

Østergaard, L., Larsen, N., Leffers, H., Kjems, J., Garret, R.: A ribosomal RNA operon and its flanking region from the archaebacterium *Methanobacterium thermoautotrophicum*, Marburg strain: transcription signals, RNA structure and evolutionary implications. System. Appl. Microbiol. *9*, 199-209 (1987)

Saitou, N., Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425 (1987)

Sambrook, J., Fritsch, E. F., Maniatis, T.: Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. (1989)

Sanger, F., Nicklen, S., Coulson, A. R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463-5467 (1977)

Shuber, A. P., Orr, E C., Recny, M. A., Schendel, P. F., May, H. D., Schauer, N. L., Ferry, J. G.: Cloning, expression, and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum*. J. Biol. Chem. *261*, 12942-12947 (1986)

Touzel, J. P., Conway de Macario, E, Nölling, J., de Vos, W. M., Zhilina, T., Lysenko, A. M.: DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol., *42*, 408-411 (1992)

Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, L. H., Moore, W. E. C., Murray, R. G. E., Stackebrandt, E., Starr, M. P., Trüper, H. G.: Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. *37*, 463-464 (1987)

Whitman, W. B., Bowen, T. L., Boone, D. R.: The methanogenic bacteria, pp. 719-767. In: The Prokaryotes, 2nd ed. (eds. **A. Balows, H. G. Trüper, M. Dworkin, W. Harder, K. H. Schleifer**), Springer-Verlag, New York (1992)

Woese, C.R.: Bacterial evolution. Microbiol. Rev. *51*, 221-271 (1987)

Yamamoto, K., Tachibana, A., Dhavises, G., Tanaka, T., Taniguchi, M., Oi, S.: Characterization of a thermophilic formate-utilizing methanogen, *Methanobacterium thermoformicicum* strain SF-4. Agricultural and Biological Chemistry *53*, 533-543 (1989)

Yanisch-Perron, C., Vieira, J., Messing, J.: Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene *33*, 103-119 (1985)

ΦF1 and ΦF3, two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*

Jörk Nölling, Alexander Groffen and Willem M. de Vos

SUMMARY

Two virulent, archaeal phages, ΦF1 and ΦF3, were isolated that were capable of infecting different thermophilic members of *Methanobacterium*. Both phages exhibited a similar morphology consisting of a polyhedral head and a tail but differed considerably in their host specificities and the size and topology of their genomes. Phage ΦF1 contained a linear, double-stranded DNA genome of 85 ± 5 kb in size and showed a broad host range including *M.thermoformicicum* strains Z-245, FTF, FF1, FF3 and CSM3, and *M.thermoautotrophicum* ΔH. In contrast, ΦF3 phage particles contained a circular genome, comprising approximately 36 ± 2 kb double-stranded DNA, and could only be propagated on *M.thermoformicicum* FF3. Hybridization experiments did not reveal similarity between the genomes of ΦF1 and ΦF3 nor between both phages and genomic DNA from different thermophilic members *Methanobacterium* or from phage ψM1 of *M.thermoautotrophicum* Marburg. A physical map of both phage genomes was constructed. The DNA of phage ΦF1 was found to contain multiple GGCC-sites which form the target for restriction-modification (R/M) system *Mth*TI of *M.thermoformicicum* THF. Therefore, restriction of ΦF1 DNA may explain the resistance of strain THF against infection by ΦF1. In contrast, the ability of ΦF1 to infect strains Z-245 and FTF both of which harbor CTAG-specific R/M systems, is probably due to the presence of only a single CTAG-site in the genome of ΦF1.

INTRODUCTION

A variety of viruses and virus-like particles have been found in representatives of the three major lines of the archaebacteria (now termed *Archaea*; Woese et al., 1990), the methanogenic, the halophilic and the extreme thermophilic, sulfur-dependent organisms (for review: Zillig et al., 1988). All archaeal viruses and virus-like particles characterized so far contain double-stranded DNA genomes with a size ranging from less than 30 kb found for those infecting extreme thermophilic, sulfur-dependent *Archaea* up to 230 kb for the phage Ja1 of *Halobacterium halobium* (Zillig et al., 1988; Wais et al., 1975).

One of the best studied archaeal phages is the lemon-shaped, temperate virus SSV1 from *Sulfolobus shibatae* whose circular, positively supercoiled 15.5-kb genome (Nadal et al., 1986) is maintained either free or integrated into the chromosome of its host (Yeats et al., 1982). The integration site (Reiter et al., 1989) and transcription units (Reiter et al., 1987) of the SSV1 genome have been mapped and its complete nucleotide sequence has been determined (Palm et al., 1991). Initially described as a virus-like particle which could not

be propagated on *S.shibatae*, SSV1 was recently shown to infect *S.solfataricus* and is thus a true virus (Schleper et al., 1992). Moreover, also transfection with SSV1 DNA has been demonstrated (Schleper et al., 1992). A temperate phage of the halophilic branch, ΦH, has also successfully been used to transfect spheroplasts of its host *Halobacterium halobium* (Cline and Doolittle, 1987). The linear 59-kb ΦH genome, whose transcription units were mapped (Zillig et al., 1986), is circularized upon lysogenization (Schnabel and Zillig, 1984b) and exhibits, similar to the genome of its host, a high degree of genetic variability (Schnabel et al., 1982) which is mediated by certain insertion elements, predominantly ISH1.8 (Zillig et al., 1988; Schnabel et al., 1984a). Similar to certain phages of the bacterial domain, modification of archaeal phage DNA has been observed for phage ΦN of *H.halobium* whose linear, 56-kb genome was found to contain 5-methyl-cytosine (Vogelsang-Wenke and Oesterhelt, 1988).

For methanogens, at least one virus-like particle, isolated from *Methanococcus voltae* A3 (Wood et al., 1989) and a putative prophage of *Methanolobus vulcani* (Wilharm et al., 1986) have been described. Two infective phage particles were isolated from *Methanobrevibacter smithii*, the lytic phages PG (Bertani and Baresi, 1986) and PMS1 (M. R. Knox and J. E. Harris, Abstr. XIV Internat. Congr. Microbiol. 1986, P.G3-8, p. 240) with genome sizes of 50 and 35 kb, respectively. The best characterized phage of the methanogenic branch is ψM1 from the thermophilic *Methanobacterium thermoautotrophicum* Marburg (Meile et al., 1989). This lytic phage contains a linear, circularly permutated 27.1-kb genome with a terminal redundancy of approximately 3 kb (Jordan et al., 1989). Interestingly, besides the concatemeric precursor of the phage DNA which is packaged by a headful mechanism, also multimers of the cryptic 4.4-kb plasmid pME2001 harbored by strain Marburg (Meile et al., 1983) could be isolated from ψM1 particles (Meile et al., 1989). This relaxed packaging machinery is probably responsible for the recently demonstrated general transduction mediated by ψM1 (Meile et al., 1990).

In the present study we report the isolation and initial characterization of two novel lytic archaeal phages, ΦF1 and ΦF3, capable of infecting different thermophilic strains of the genus *Methanobacterium*.

MATERIAL AND METHODS

Bacterial strains and phages

Strains of *M.thermoformicicum* and *M.thermoautotrophicum* used in this study are listed in Table 1. Cultivation of methanogenic strains in liquid cultures or on solidified 0.8% Gelrite plates (Roth GmbH, Karlsruhe, FRG) was performed as described previously (Nölling et al., 1991) except that 0.2 g/l $MgCl_2$ was used. DNA of phage ψM1 was a gift from L. Meile (Department of Microbiology, ETH Zürich, Switzerland)

Isolation of methanogenic phages

For the isolation of both phages ΦF1 and ΦF3 samples were taken anaerobically from an upflow anaerobic sludge-bed reactor operating at 55°C (Department of Environmental Engineering, University of Wageningen, The Netherlands) which already had served as source for the isolation of *M.thermoformicicum* strains FF1 and FF3 (Nölling et al., 1991).

For enrichment of phages, 10 ml of the samples were sterilized by filtration on a 0.2 μm filter (Schleicher & Schuell, Dassel, FRG) and added to 100 ml serum bottles containing 20 ml medium described by Nölling et al. (1991). These serum bottles were inoculated with 1 ml of exponentially grown cultures of either strain FF1 or FF3 and incubated for at least two days under H_2/CO_2 at 55°C. Samples of the grown enrichment culture were tested for plaque formation on a lawn of the indicator strains FF1 and FF3.

Table 1. Methanogenic strains used in this study

Strain	Plasmid (size in kb)	Reference
M.thermoformicicum		
Z-245 (DSM 3720)*	pFZ1 (11)	(Zhilina et al., 1984)
FTF (DSM 3012)	pFZ2 (11)	(Touzel et al., 1988)
THF (DSM 3848)	pFV1 (13.5)	(Zinder and Koch, 1984)
CB12 (DSM 3664)	-	(Zhao et al., 1984)
SF-4	-	(Yamamoto et al., 1989)
FF1	-	(Nölling et al., 1991)
FF3	-	(Nölling et al., 1991)
CSM3	-	(Nölling et al., 1993)
HN4	-	(Nölling et al., 1993)
M.thermoautotrophicum		
ΔH (DSM 1053)	-	(Zeikus and Wolfe, 1972)
Marburg (DSM 2133)	pME2001 (4.4)	(Fuchs et al., 1978)

*DSM, German Collection of Microorganisms, Braunschweig, FRG.

Propagation and titration of phages

For small scale preparation of phage lysates, a single plaque was added to 100 ml serum bottles containing 20 ml medium (Nölling et al., 1991) inoculated with 1 ml of exponentially grown cultures of the appropriate methanogenic host (about 5 x 10^7 cells/ml) and incubated for one or two days at 55°C under H_2/CO_2 atmosphere in a shaking incubator at 150 rpm. Large quantities of phage lysate were obtained under similar conditions using 1 l serum bottles containing 0.2 l medium inoculated with 0.2 ml phage lysate (approximately 10^6 PFU/ml) and 1 ml of a methanogenic culture. Alternatively, high titer lysates (10^9 PFU/ml) were prepared from plate lysates obtained by flooding agar plates showing confluent lysis with 1.5 ml SM-buffer (Sambrook et al., 1989). For routine propagation of phages, phage lysates were prepared in 100 ml serum bottles as described above, sterilized by filtration through a 0.2 μm filter and stored anaerobically at 4°C.

Phage titers were determined by the soft agar overlay method essentially as described by Meile et al. (1989). Usually, 0.1 ml of (diluted) phage lysate was mixed with 0.3 ml exponentially growing cells and incubated for 20 min at room temperature to allow phage adsorption. Soft agar (2.5 ml of 0.7% agar) was added and the mixture was poured onto

freshly prepared and dried Gelrite plates (Nölling et al., 1991). The plates were incubated upside down at 55°C under H_2/CO_2 atmosphere for one to three days.

Isolation of DNA

Phage lysates were centrifugated at 10,000 x g for 10 min and the supernatant was collected. DNase and RNase A (Boehringer GmbH, Mannheim, FRG) were added to a final concentration of 0.2 μg/ml, and the supernatant was gently stirred for 1 h at room temperature. Phage particles were precipitated by adjusting the supernatant to 1 M NaCl and 10% (wt/vol) polyethylene glycol 6000, followed by an overnight incubation at 4°C. The phage precipitate was recovered by centrifugation at 10,000 x g for 10 min at 4°C and gently resuspended in SM-buffer. After addition of sodium dodecyl sulfate and EDTA to a final concentration of 0.5% and 20 mM, respectively, the suspension was treated with proteinase K (0.1 mg/ml; Boehringer GmbH, Mannheim, FRG) for 1 h at 56°C. The DNA was deproteinized by several phenol/chloroform extractions followed by an extraction with chloroform. Finally, the DNA was precipitated with ethanol and dissolved in TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA).

Total DNA from strains of the genus *Methanobacterium* was isolated as described by Nölling et al. (1991).

Manipulation and hybridization of DNA

All restriction enzymes were purchased from Life Technologies Inc. (Gaithersburg, Md.) and used according to the instructions of the supplier. R.*Mth*ZI and R.*Mth*TI endonuclease assays were performed as described earlier using extracts prepared from *M.thermoformicicum* Z-245 (Nölling and de Vos, 1992b) and from *E.coli* harboring pURM4ΔM with the cloned *mthTIR* gene from plasmid pFV1 of *M.thermoformicicum* THF (Nölling and de Vos, 1992a), respectively. To remove histone-like proteins present in the extracts prepared from strain Z-245, R.*Mth*ZI assays were deproteinized prior gel electrophoresis as described previously (Nölling and de Vos, 1992b).

For DNA-DNA hybridization, DNA was transferred from agarose gels to Hyperbond membranes by a capillary blot procedure as specified by the supplier (Amersham, United Kingdom). For preparation of DNA probes, either total phage DNA or DNA fragments isolated from agarose gels using the Geneclean kit (Bio 101, La Jolla, Calif.) were labelled by nick-translation (Sambrook et al., 1989) with [α-^{32}P]dATP (Amersham, United Kingdom). Prehybridization and hybridization were performed as described earlier (Nölling et al., 1991). Washing procedures were carried out in 2 x SSC/0.1% SDS at 55°C (medium stringent) and in 0.1 x SSC/0.1% SDS at 65°C (stringent).

Electron microscopy

Preparations of ΦF1 and ΦF3 were placed on Formvar-coated cooper grids and stained with 1% uranyl acetate. The electron microscope used for examination was a Philips CM12.

RESULTS

Morphology of phage ΦF1 and ΦF3

We have previously reported the isolation of the *M.thermoformicicum* strains FF1 and FF3 from an anaerobic, thermophilic digestor (Nölling et al., 1991). Both strains served as indicator for the enrichment and subsequent isolation of two phages from the same anaerobic digestor, which accordingly were designated ΦF1 and ΦF3.

Electron microscopy of negatively stained phage preparations revealed a similar morphology of both phages ΦF1 and ΦF3. Phage ΦF1 consists of a isometric, icosahedric capsid approximately 70 nm in diameter and a nonflexible tail of approximately 160 nm in length and 20 nm wide (Fig. 1A). Phage ΦF3 is composed of a isometric, icosahedric head of approximately 55 nm in diameter and a 230 nm long and 9 nm wide tail which is flexible in contrast to that of phage ΦF1 (Fig. 1B). The ΦF3 tail consists of about 60 transversal segments (Fig. 1B). Since no shortened tails were detected, the tails of ΦF1 and ΦF3 are probably not contractible. The morphology of both ΦF1 and ΦF3 is that of Bradley's classification group B (Bradley, 1967).

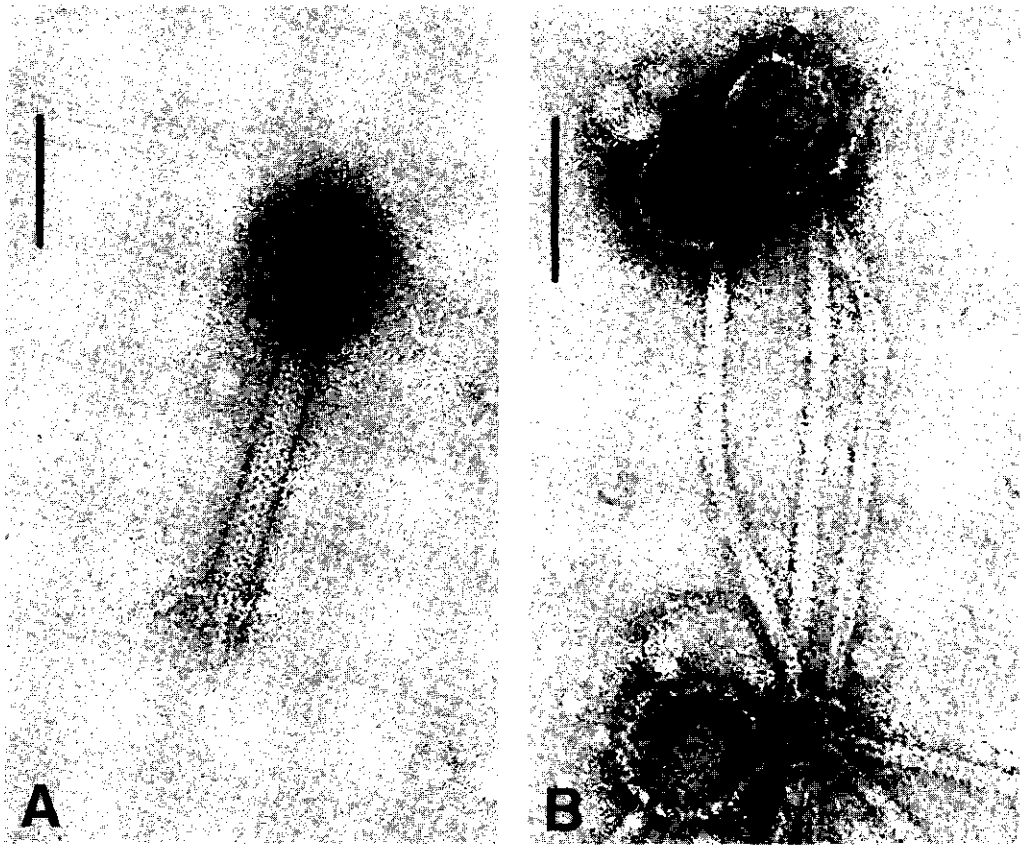


Figure 1. Negatively stained ΦF1 (A) and ΦF3 (B) particles. The bars represent 70 nm.

Host specificity of phage ΦF1 and ΦF3

All known strains of *M.thermoformicicum* and the *M.thermoautotrophicum* strains Marburg and ΔH were tested for sensitivity to phages ΦF1 and ΦF3. As summarized in Table 2 both phages differ with respect to their host range. While phage ΦF3 could only be propagated on *M.thermoformicicum* FF3, phage ΦF1 was found to infect a variety of strains including *M.thermoformicicum* strains Z-245, FTF, FF1, FF3 and CSM3, and *M.thermoautotrophicum* ΔH. No plaque formation was observed when each of the other strains was used as an indicator even not after increasing the number of phages to approximately 10^5 PFU per plate.

Table 2. Host range of phage ΦF1 and ΦF3

Strain		ΦF1[a]	ΦF3[a]
M.thermoformicicum	Z-245	+	-
	FTF	+	-
	THF	-	-
	CSM3	+	-
	FF1	+	-
	FF3	+	+
	CB12	-	-
	SF-4	-	-
	HN4	-	-
M.thermoautotrophicum	ΔH	+	-
	Marburg	-	-

[a] +, indicates efficiency of plating of more than 10^{-1} while
-, indicates efficiency of plating of less than 10^{-5}.

Lysates produced by ΦF1 on either host and ΦF3 lysates contained similar amounts of infective particles and titers of 1-5 x 10^6 PFU/ml were obtained. Phage ΦF1 formed clear and small (< 1 mm in diameter) plaques that became visible after 2-3 days of incubation. The clear plaques produced by phage ΦF3 had a size of approximately 2 mm in diameter and could already be detected after about 20 h of incubation. Storing the lysates produced by either phage in the presence of oxygen did not affect the phage titer.

Genomes of phage ΦF1 and ΦF3

The genome of the phages ΦF1 and ΦF3 is a double-stranded DNA molecule which served as a good substrate for cleavage by seventeen type II restriction endonucleases with different specificity. For the calculation of the genome sizes and for the construction of physical maps, we have used several commercially available restriction endonucleases and those produced by *M.thermoformicicum* strains THF and Z-245, the GGCC-specific R.*Mth*TI (Nölling and de Vos, 1989a) and the CTAG-recognizing R.*Mth*ZI (Nölling and de Vos, 1989b), respectively.

The size of the ΦF1 genome was estimated to be 85 ± 5 kb by summing up the length of

the restriction fragments derived from digestion with *Sal*I (15 fragments; Fig. 2) and a triple digest with *Xba*I-*Not*I-*Mlu*I (eight fragments). Further restriction analysis revealed no sites for *Sfi*I, one site for *Xba*I and *Mth*ZI (Fig. 2) and three sites for *Mlu*I (Fig. 2) and *Not*I. The latter three enzymes were used in double- and triple digestions for the construction of a physical map which showed that ΦF1 contains a linear genome (Fig. 3). Incubation of ΦF1 DNA with *Mth*TI generated multiple fragments, the majority of which were smaller than 1 kb (data not shown). To examine whether ΦF1 contains cohesive ends, *Mlu*I-digested phage DNA was heated to 78°C for 10 min and immediately cooled on ice. Separation by agarose gel electrophoresis together with non-heated *Mlu*I-digested ΦF1 DNA did not reveal any difference in restriction pattern (data not shown).

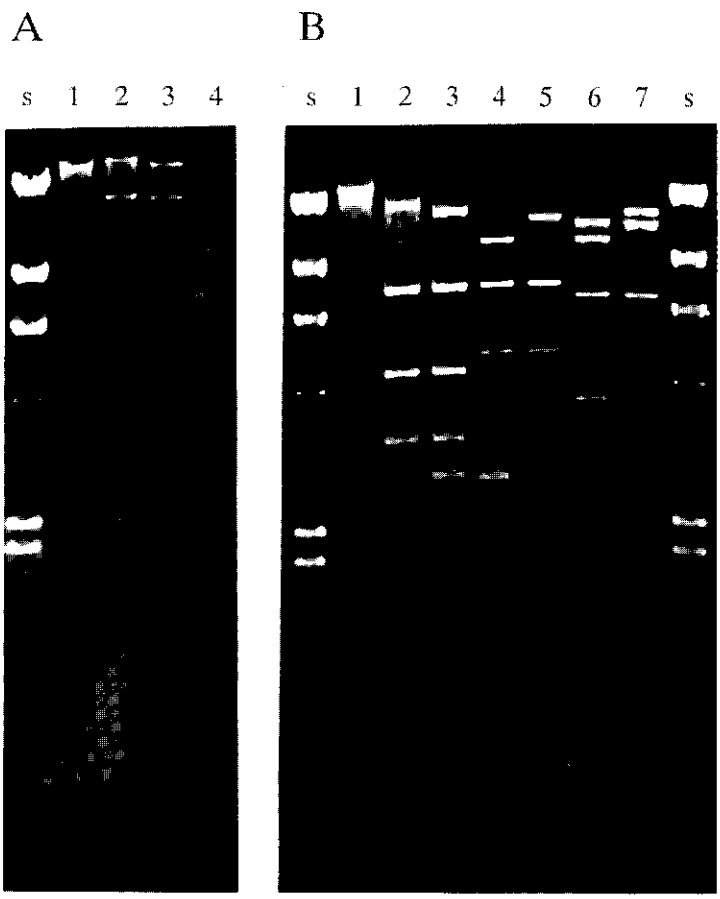


Figure 2. Restriction analysis of DNA prepared from phage ΦF1 (panel A) and ΦF3 (panel B). (A) Approximately 0.3 μg ΦF1 DNA undigested (lane 1), and digested with *Mth*ZI (lane 2), *Mlu*I (lane 3) and *Sal*I (lane 4). (B) Approximately 0.5 μg ΦF3 DNA undigested (lane 1), and digested with *Sal*I (lane 2), *Sal*I-*Sst*I (lane 3), *Sal*I-*Sst*I-*Mth*ZI (lane 4), *Sal*I-*Mth*ZI (lane 5), *Sst*I-*Mth*ZI (lane 6), and *Mth*ZI (lane 7). Lanes s show *Hind*III-digested λ DNA as size marker.

Restriction digestion of phage ΦF3 DNA with *Sal*I and *Mth*ZI produced four and three fragments, respectively (Fig. 2). If the ΦF3 genome is linear, this would correspond to three *Sal*I- and two *Mth*ZI-sites for the ΦF3 genome which would generate six restriction fragments upon a double digestion with both endonucleases. However, we observed seven fragments derived from digestion of ΦF3 DNA with *Sal*I followed by incubation with *Mth*ZI (Fig. 2).

Interpretation of these results led to the conclusion that phage ΦF3 contains a circular genome. Mapping of the restriction sites for *Sst*I, *Hind*III, *Sal*I and *Mth*ZI confirmed the circular configuration of the ΦF3 genome which had a calculated size of 36 ± 2 kb (Fig. 3). The circularization of the ΦF3 genome seems not to be mediated by cohesive ends since heated (78°C for 10min) and non-heated *Sal*I-digests of ΦF3 DNA showed no differences in restriction pattern after separation by agarose gel electrophoresis (data not shown).

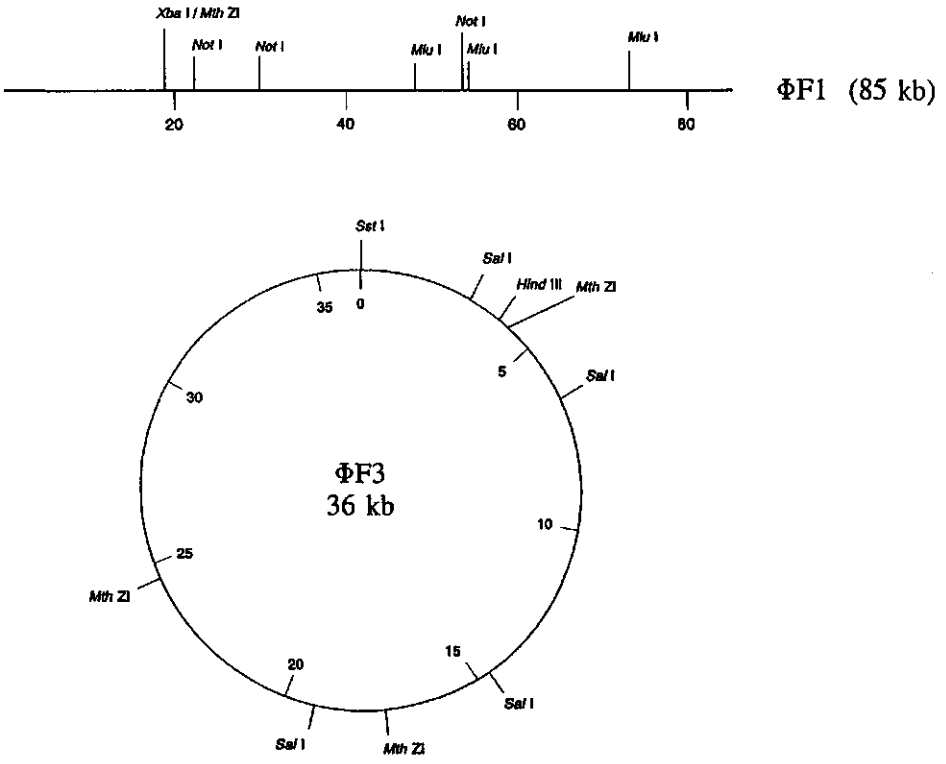


Figure 3. Physical maps of the linear genome of ΦF1 and the circular genome of ΦF3. Sizes are given in kilobases.

Homology of ΦF1 and ΦF3 DNA with DNA from other sources

Hybridization studies using radioactively labelled ΦF1- and ΦF3 DNA revealed no similarity to digested total DNA from the *Methanobacterium* strains used in this study (data not shown). In addition, no cross-hybridization was observed between DNA from ΦF1, ΦF3, ψM1 DNA (data not shown).

Since phage ψM1 was found to package multimers of plasmid pME2001 harbored by its host *M.thermoautotrophicum* Marburg (Meile et al., 1989), we tested whether phage ΦF1 contained (fragments of) plasmid pFZ1 DNA from *M.thermoformicicum* Z-245. For this purpose, we extracted DNA from phages propagated on strain Z-245 and the plasmid-free

M.thermoautotrophicum ΔH and hybridized it, after digestion and subsequent blotting, to the three cloned *Kpn*I fragments of plasmid pFZ1 labelled by nick-translation (Nölling et al., 1991). The DNA from ΦF1 phage particles propagated either on strain showed no hybridization with pFZ1 (data not shown) indicating that ΦF1 does not have the capacity to package pFZ1 DNA.

DISCUSSION

Two novel lytic phages ΦF1 and ΦF3 capable of infecting different thermophilic strains of the archaeal genus *Methanobacterium* were described. Both phages ΦF1 and ΦF3 were composed of a polyhedral head and a tail similar to the morphology of the phages found for halophilic and methanogenic *Archaea*, with the exception of the virus-like particle of *Methanococcus voltae* A3 (Wood et al., 1989).

Neither ΦF1 nor ΦF3 contained sequences with similarity to DNA from any of the phages or methanogens used in this study. The lack of sequence similarity between ΦF1, ΦF3 and ψM1 would imply that these phages use different mechanism to induce lysis of the host cell. Since lysis of *M.thermoautotrophicum* Marburg is mediated by a phage encoded pseudomurein endopeptidase (Stax et al., 1992), ΦF1 and ΦF3 may either specify a different lytic activity or induce autolysis of the host.

Phage ΦF1 contains a linear genome consisting of approximately 85 kb which so far is the largest genome reported for a methanogenic phage. In contrast, DNA extracted from ΦF3 phage particles was approximately 36 kb in size and showed a circular configuration which is an unusual topology for the genome of a tailed phage. A similar genome topology was, however, found for the virus-like particle from *Methanococcus voltae* A3 which contains a 23-kb DNA molecule (Wood et al., 1989) and the 15.5-kb genome of the virus SSV1 from *Sulfolobus shibatae* (Nadal et al., 1986). The circular form of the ΦF3 DNA suggests that its genome is either replicated and packaged as a monomer or that replication of ΦF3 DNA results in a (linear or circular) multimeric form which is resolved into circular monomers during packaging.

The phages ΦF1 and ΦF3 differ remarkably in their host range. Similar to phage ψM1 of *M.thermoautotrophicum* Marburg, phage ΦF3 has a narrow host range and could only be propagated on *M.thermoformicicum* FF3 which served as indicator strain for the isolation of ΦF3. In contrast, phage ΦF1 was capable to infect five strains of *M.thermoformicicum* and *M.thermoautotrophicum* ΔH. This broad host-range of ΦF1 is relevant in the context of the recently reported phylogenetic studies on thermophilic strains of the genus *Methanobacterium*. Based on DNA-DNA similarity (Touzel et al., 1992) and 16S rRNA analysis (Nölling et al., 1993) these studies provided evidence for the conclusion that (i) the species *M.thermoformicicum* can be divided into two groups of strains, the Z-245- and CB12-group, which may constitute different species, (ii) strain *M.thermoautotrophicum* ΔH belongs to the Z-245-group, and (iii) *M.thermoautotrophicum* Marburg is clearly different from either of those strains. The host range of ΦF1 that includes *M.thermoformicicum* strains Z-245, FTF, CSM3, FF1 and FF3, and *M.thermoautotrophicum* ΔH, all of which are closely related members of the Z-245-group, supports the proposed different taxonomic position of these thermophilic *Methanobacterium* strains (Touzel et al., 1992; Nölling et al., 1993).

Phage ΦF1 failed, however, to infect *M.thermoformicicum* THF which is a more distantly related member of the Z-245-group (Touzel et al., 1992; Nölling et al., 1993). Although

strain THF could be a trivial phage-resistant mutant it is also possible that the penetrated ΦF1 DNA is inactivated by endonuclease attack of the plasmid-encoded, GGCC-recognizing type II restriction-modification (R/M) system *Mth*TI of strain THF (Nölling and de Vos, 1992a). If so, the R/M system of strain THF would be more efficient in protecting against phage ΦF1 infection than the R/M systems *Mth*ZI and *Mth*FI of strains Z-245 and FTF, respectively, that are specific for the sequence CTAG (Nölling and de Vos, 1992b). The extreme difference in the frequency of appropriate endonuclease target sites in the 80 kb genome of ΦF1, i.e. a single CTAG-site versus multiple GGCC-sites, would support this assumption (Fig. 2 and 3). Although CTAG is one of the rarest occurring tetranucleotides in DNA of methanogens (Nölling, unpublished results) the exceptionally strong bias against CTAG found in the genome of ΦF1 could well be a result of counterselection against restriction sites in the phage DNA in order to escape the host-controlled restriction barrier. Such an antirestriction mechanism is one of the phage-derived defense strategies known to be realized by certain bacterial bacteriophages to avoid restriction by the host (Krüger and Bickle, 1983).

The specificity of phage ΦF3 for *M.thermoformicicum* FF3 as only host was unexpected since strain FF3 belongs to the group of closely related strains that serve as host for ΦF1. The existence of a type II restriction barrier is not likely since, in contrast to *M.thermoformicicum* strains THF, Z-245 and FTF, extracts prepared from strains FF3, FF1 and CSM3 and *M.thermoautotrophicum* ΔH did not contain detectable site-specific endonuclease activities (Nölling and de Vos, 1992b). These findings suggest that phage ΦF3 depends on a specific factor which is only present in strain FF3 or that the other strains of the Z-245-group contain additional phage defense mechanisms such as type I or III R/M systems.

ACKNOWLEDGEMENTS

We are very grateful to J. Groenewegen (Department of Virology, Wageningen Agriclutural University) for his technical assistance in electron microscopy.

REFERENCES

Bertani, G. & Baresi, L. (1986). Looking for a gene transfer mechanism in methanogenic bacteria, in: Archaebacteria'85 (O. Kandler and W. Zillig, eds.), Gustav Fischer Verlag, Stuttgart.

Bradley, D. E. (1967). Ultrastructure of bacteriophages and bacteriocins. Bacteriological Reviews **31**, 230-314.

Cline, S. W. & Doolittle, W. F.. (1987). Efficient transfection of the archaebacterium *Halobacterium halobium*. Journal of Bacteriology **169**, 1341-1344.

Fuchs, G., Stupperich, E. & Thauer, R. K. (1978). Acetyl CoA, a central intermediate of autotrophic CO_2 fixation in *Methanobacterium thermoautotrophicum*. Archives of Microbiology **127**, 267-272.

Gibson, T. J. (1984). Studies on the Epstein-Barr virus genome. Ph.D. thesis. Cambridge University, Cambridge, England.

Jordan, M., Meile, L. & Leisinger, T. (1989). Organization of *Methanobacterium thermoautotrophicum* bacteriophage ψM1 DNA. Molecular and General Genetics **220**, 161-164.

Krüger, D. H. & Bickle, T. A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiological Reviews **47**, 345-360.

Meile, L., Kiener, A. & Leisinger, T. (1983). A plasmid in the archaebacterium *Methanobacterium*

thermoautotrophicum. Molecular and General Genetics **191**, 480-484.

Meile, L., Jenal, U., Studer, D., Jordan, M. & Leisinger, T. (1989). Characterization of ψM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. Archives of Microbiology **152**, 105-110.

Meile, L., Abendschein, P. & Leisinger, T. (1990). Transduction in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. Journal of Bacteriology **172**, 3507-3508.

Nadal, M., Mirambeau, G., Forterre, P., Reiter, W.-D. & Duguet, M. (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. Nature (London) **321**, 256-258.

Nölling, J., Frijlink, M. & de Vos, W. M. (1991). Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. Journal of General Microbiology **137**, 1981-1986 (Chapter 4).

Nölling, J. & de Vos, W. M. (1992a). Characterization of the archaeal, plasmid-encoded type II restriction-modification system *Mth*TI from *Methanobacterium thermoformicicum* THF: homology to the bacterial *Ngo*PII system from *Neisseria gonorrhoeae*. Journal of Bacteriology **174**, 5719-5726 (Chapter 7).

Nölling, J. & de Vos, W. M. (1992b). Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene. Nucleic Acids Research **20**, 5047-5052 (Chapter 8).

Nölling, J., Hahn, D., Ludwig, W. & de Vos, W. M. (1993). Phylogenetic analysis of thermophilic *Methanobacterium* sp.: evidence for a common formate-utilizing ancestor. Systematic and Applied Microbiology, in press (Chapter 2).

Palm, P., Schleper, C., Grampp, B., Yeats, S. McWilliam, P., Reiter, W.-D. & Zillig, W. (1991). Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185**, 242-250.

Reiter, W.-D., Palm, P. Yeats, S. & Zillig, W. (1987). Gene expression in archaebacteria: Physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. Molecular and General Genetics **209**, 270-275.

Reiter, W.-D., Palm, P. & Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Research **17**, 1907-1914.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Schleper, C., Kubo, K. & Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. Proceedings of the National Academy of Sciences of the United States of America **89**, 7645-7649.

Schnabel, H., Schramm, E., Schnabel, R. & Zillig, W. (1982). Structural variability in the genome of ΦH of *Halobacterium halobium*. Molecular and General Genetics **193**, 370-377.

Schnabel, H., Palm, P., Dick, K. & Grampp, B. (1984a). Sequence analysis of the insertion element ISH1.8 and of associated structural changes in the genome of phage ΦH of the archaebacterium *Halobacterium halobium*. EMBO Journal **3**:1717-1722.

Schnabel, H. & Zillig, W. (1984b). Circular structure of the genome of phage ΦH in a lysogenic *Halobacterium halobium*. Molecular and General Genetics **193**, 422-426.

Stax, D., Hermann, R., Falchetto, R. & Leisinger, T. (1992). The lytic enzyme in bacteriophage ψM1-induced lysates of *Methanobacterium thermoautotrophicum* Marburg. FEMS Microbiology Letters **100**, 433-438.

Touzel, J. P., Petroff, D., Maestrojuan, G. M., Prensier, G. & Albagnac, C. (1988). Isolation and characterization of a thermophilic *Methanobacterium* able to use formate, the strain FTF. Archives of Microbiology **149**, 291-296.

Touzel, J. P., Conway de Macario, E., Nölling, J., de Vos, W. M., Zhilina, T. & Lysenko, A. M. (1992). DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methano-*

bacterium thermoformicicum as a synonym of *Methanobacterium thermoautotrophicum*. International Journal of Systematic Bacteriology **42**, 408-411.

Vogelsang-Wenke, H., & Oesterhelt, D. (1988). Isolation of a halobacterial phage with a fully cytosine-methylated genome. Molecular and General Genetics **211**, 407-414.

Wais, A. C., Kon, M., MacDonald, R. E. & Stollar, B. D. (1975). Salt-dependent bacteriophage infecting *Halobacterium cutirubrum* and *H.halobium*. Nature (London) **256**, 314-315.

Wilharm, T., Thomm, M. & Stetter, K. O. (1986). Genetic analysis of plasmid pMP1 from *Methanolobus vulcani*. Systematic and Applied Microbiology **7**, 401.

Wood, A. G., Whitman, W. B. & Konisky, J. (1989). Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. Journal of Bacteriology **171**, 93-98.

Yamamoto, K., Tachibana, A., Dhavises, G., Tanaka, T., Taniguchi, M. & Oi, S. (1989). Characterization of a thermophilic formate-utilizing methanogen, *Methanobacterium thermoformicicum* strain SF-4. Agricultural and Biological Chemistry **53**, 533-543.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**, 103-119.

Yeats, S., McWilliam, P. & Zillig, W. (1982). A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO Journal **1**, 1035-1038.

Zhao, Y., Zhang, H., Boone, D. & Mah, R. A. (1986). Isolation and characterization of a fast-growing, thermophilic *Methanobacterium* species. Applied and Environmental Microbiology **52**, 1227-1229.

Zeikus, J. G. & Wolfe, R. S. (1972). *Methanobacterium thermoautotrophicum* sp. n., an anaerobic, autotrophic, extreme thermophile. Journal of Bacteriology **109**:707-713.

Zhilina, T. N. & Ilarionov, S. A. (1984). Characteristics of formate assimilating methane bacteria and description of *Methanobacterium thermoformicicum* sp. nov. Microbiology **53**, 647-651 (English translation of Mikrobiologyia **53**, 785-790).

Zillig, W., Reiter, W.-D., Palm, P., Gropp, F., Neumann, H. & Rettenberger, M. (1988). Viruses of archaebacteria. In: The Bacteriophages, Vol. 1, p. 517-558 (eds. R. Calendar), Plenum Press, New York.

Zinder, S. & Koch, M. (1984). Non-acetoclastic methanogenesis from acetate: acetate oxidation by a thermophilic syntrophic coculture. Archives of Microbiology **138**, 263-272.

Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*

Jörk Nölling, Maarten Frijlink, and Willem M. de Vos

Journal of General Microbiology **137**, 1981-1986.

Journal of General Microbiology (1991), **137**, 1981–1986. *Printed in Great Britain*

Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*

JÖRK NÖLLING, MAARTEN FRIJLINK and WILLEM M. DE VOS*

Bacterial Genetics Group, Department of Microbiology, Agricultural University of Wageningen, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

(*Received 6 December 1990; revised 22 March 1991; accepted 29 April 1991*)

Seven strains of the formate-utilizing thermophilic methanogen *Methanobacterium thermoformicicum* were screened for the presence of extrachromosomal DNA. Covalently closed circular plasmid DNA was detected in three strains, Z-245, FTF and THF. No plasmids were found in strains CB12 and SF-4, nor in two new isolates, FF1 and FF3. The plasmid from strain Z-245, designated pFZ1, and that from strain FTF, designated pFZ2, were approximately 10·5 kb in size and contained homologous and similarly sized restriction fragments. A physical map of plasmid pFZ1 was constructed and three *Kpn*I-fragments comprising the entire plasmid were cloned in *Escherichia coli* using pUC19. Plasmid pFV1 isolated from strain THF was approximately 14 kb in size and contained regions with strong homology to pFZ1 DNA. Using the *Methanobacterium formicicum fdhA* gene as a hybridization probe the strains of the species *M. thermoformicicum* could be classified into two major groups.

Introduction

The concept of the Archae forming one of the three domains of living organisms (Woese *et al.*, 1990) has generated considerable interest in identifying and characterizing archaeal extrachromosomal elements in order to study gene transfer in and develop transformation systems for these unique organisms. Furthermore, detailed knowledge of these extrachromosomal elements contributes to understanding their evolution and their possible interaction with the chromosome of their hosts.

The presence of viruses and virus-like particles has been described in various representatives of the archaeal domain, including sulphur-dependent, halophilic and methanogenic species (for reviews, see Brown *et al.*, 1989; Wood *et al.*, 1989). Several plasmids up to 150 kb in size have been isolated from halophilic bacteria (Brown *et al.*, 1989). Some of these appear to be involved in genome instability (Pfeifer *et al.*, 1981) and carry genes with known functions (DasSarma *et al.*, 1987). In contrast, all plasmids isolated so far from methanogens are cryptic. These include four plasmids derived from *Methanococcus* (Wood *et al.*, 1985; Zhao *et al.*, 1988), one from *Methanosarcina acetivorans* (Sowers & Gunsalus, 1988), one from *Methanolobus vulcani* (Thomm *et al.*, 1983), which also may be a phage (Wilharm *et al.*, 1986), and a plasmid from *Methanobacterium thermoautotrophicum* strain Marburg (Meile *et al.*, 1983). In this report we describe the isolation of plasmid DNA from three strains of *M. thermoformicicum* and the characterization of one of these plasmids, the cryptic plasmid pFZ1 from strain Z-245. Furthermore, we identify plasmid-free *M. thermoformicicum* strains which may be used as recipient strains in a transformation system.

Methods

Bacterial strains and vectors. Table 1 lists the strains of *M. thermoformicicum* used in this study. *M. formicicum* strain MS1 (DSM 3636) was kindly provided by G. D. Vogels (Department of Microbiology, University of Nijmegen, The Netherlands). *M. wolfei* (DSM 2970) and *M. thermoautotrophicum* strains Marburg (DSM 2133) and ΔH (DSM 1053) were from the German Collection of Microorganisms, Braunschweig, FRG. *Escherichia coli* strain DH5α and plasmid pUC19 (Yanisch-Perron *et al.*, 1985) were obtained from Life Technologies.

Growth conditions. M. formicicum, M. thermoformicicum and *M. thermoautotrophicum* were grown on minimal medium containing, per litre (J. P. Touzel, personal communication): KH_2PO_4, 0·3 g; NH_4Cl, 1·0 g; NaCl, 0·6 g; $MgCl_2.6H_2O$, 0·1 g; $CaCl_2.2H_2O$, 0·06 g; $NaHCO_3$, 4·0 g; trace element solution, 10 ml [prepared according to Touzel & Albagnac (1983), except that 0·05 g $AlCl_3.6H_2O$ l^{-1} was added]. For cultivation in serum bottles 0·5 g cysteine hydrochloride l^{-1} was added. The pH of the medium was adjusted to 7·2. For growth of *M. wolfei*, the medium was supplemented with 2·6 mg $Na_2WO_4.2H_2O$ l^{-1} (Winter *et al.*, 1984). After autoclaving under N_2/CO_2 (4:1, v/v) the headspace was flushed with 300 kPa H_2/CO_2 (4:1) and 0·5 g $Na_2S.9H_2O$ l^{-1} was added. Except for *M. formicicum*, all methano-

Table 1. *M. thermoformicicum strains screened for the presence of plasmid DNA*

Strain	Plasmid DNA (size in kb)	Source*
Z-245 (DSM 3720)	pFZ1 (10·5)	DSM*
FTF (DSM 3012)	pFZ2 (10·5)	DSM
THF (DSM 3848)	pFV1 (14)	DSM
CB12 (DSM 3664)	-	DSM
SF-4	-	K. Yamamoto, Osaka City University, Japan
FF1	-	This work
FF3	-	This work

* DSM, German Collection of Microorganisms, Braunschweig, FRG.

genic strains were cultivated according to Schönheit *et al.* (1980) in 0·8 l and 10 l fermenters containing 0·5 l and 8 l of medium, respectively. For growth in fermenters Na_2S was replaced by 0·62 g thiosulphate l^{-1} and 0·5 g cysteine l^{-1} (J. Korteland, personal communication). In addition 10 μg streptomycin ml^{-1} was added. Prior to entering the vessel, the gas mixture was led through the gas phase of a closed bottle filled with 100 ml Na_2S (0·1 g l^{-1}) serving as a reducing agent. The cultivation temperature was 37 °C for *M. formicicum*, 55 °C for *M. thermoformicicum* strains Z-245, FTF, CB12, SF-4, FF1 and FF3, and 65 °C for *M. thermoformicicum* strain THF, *M. thermoautotrophicum* strains Marburg and ΔH and *M. wolfei*. *E. coli* strain DH5α was grown in L-broth and handled as described by Maniatis *et al.* (1982).

Isolation of DNA. Pelleted cells were resuspended in TE buffer (10 mM-Tris/HCl, pH 8·0, 1 mM-EDTA) at a concentration of 1 g wet weight in 5 ml buffer. After addition of SDS to a final concentration of 0·5%, proteinase K (0·1 mg ml^{-1}; Boehringer Mannheim) and about a one-third volume equivalent of glass beads (0·3 mm), the suspension was incubated for 1 h at 56 °C. After cooling on ice the suspension was vigorously shaken for 15 s on a Vortex mixer and then kept on ice for 15 s. This shaking/cooling procedure was repeated 5–10 times until the cell suspension became viscous. Subsequently, the suspension was thoroughly deproteinized by several phenol/chloroform extractions followed by an extraction with chloroform. Finally, the nucleic acids were precipitated with ethanol and dissolved in TE buffer. In some cases, plasmid DNA was isolated by CsCl/ethidium bromide density-gradient centrifugation (Maniatis *et al.*, 1982).

For copy-number determination of plasmid pFZ1, total DNA was isolated from lysed protoplasts of *M. thermoformicicum* strain Z-245 to avoid possible nicking. For this purpose late-exponential phase cells, grown in serum bottles as described above, were harvested and incubated anaerobically with an autolysate of *M. wolfei* until protoplasts were formed, essentially as described by Mountfort *et al.* (1986). These protoplasts were washed with phosphate/sucrose buffer (Mountfort *et al.*, 1986) and resuspended in ice-cold TE buffer to lyse the protoplasts. The DNA was further purified as described above.

Isolation of formate-utilizing, thermophilic methanogens. Samples were taken from a thermophilic (55 °C) upflow anaerobic sludge-bed reactor (Department of Environmental Engineering, University of Wageningen, The Netherlands) and inoculated into serum bottles containing 20 ml of a selective medium for formate-utilizing methanogens. This consisted of minimal medium as described above with the addition of 50 mM-sodium formate, 50 mM-Tris/HCl, pH 7·3, 0·01 g yeast extract l^{-1} and 10 ml vitamin solution l^{-1} [prepared according to Balch *et al.* (1979), except that lipoic acid was omitted]. The cultures were incubated anaerobically at 55 °C under an N_2/CO_2 atmosphere (200 kPa overpressure). When the cultures became turbid, samples were plated on selective medium supplemented with 0·62 g thiosulphate l^{-1} and 0·5 g cysteine l^{-1} instead of Na_2S and solidified with 0·8% Gelrite (Roth), followed by incubation at 55 °C under an N_2/CO_2 atmosphere. Colonies became visible after about 1 week and were transferred anaerobically into serum bottles containing selective medium followed by incubation at 55 °C. After growth had developed the cultures were inspected microscopically and methane production was determined with a Becker gas-chromatograph model 417 with a TCD detector. The plating and inoculation procedures were repeated twice for methane-producing cultures.

DNA manipulations. Chromosomal and plasmid DNA were digested with endonucleases obtained from Life Technologies. For cloning into *E. coli* (Maniatis *et al.*, 1982), DNA fragments of both vector and insert DNA were isolated from agarose gels using a Geneclean kit as described by the supplier (Bio 101). Transfer of DNA from agarose gels to GeneScreen Plus membranes (NEN–Du Pont) was carried out using a capillary blot procedure as described by the supplier. The DNA probe was produced by labelling DNA by nick-translation (Maniatis *et al.*, 1982) with [α^{32}P]dATP (Amersham). If mixed probes were used, equal amounts of DNA fragments were labelled.

Results

Isolation of plasmid DNA and characterization of plasmid pFZ1

Bacteria belonging to the species *Methanobacterium* have a rigid cell wall, with pseudomurein as the primary component, which is resistant to methods commonly used for lysis of bacterial cells. Therefore, we disrupted the cells by mechanical force using glass beads as has previously been described for the successful disruption of fungal spores (Van Etten & Freer, 1978). Both chromosomal and plasmid DNA from *M. thermoautotrophicum* strain Marburg (Meile *et al.*, 1983) could be isolated efficiently, which made the method suitable for the rapid screening of methanogens for the presence of plasmid DNA.

Using this procedure, we detected extrachromosomal DNA in *M. thermoformicicum* strain Z-245 (Fig. 1*a*, lane 3). To investigate the identity of this DNA, total DNA from strain Z-245 was subjected to CsCl/ethidium bromide density-gradient centrifugation and the covalently closed circular (ccc)-fraction, consisting of supercoiled plasmid, was isolated. Except for some residual chromosomal DNA, the DNA of the ccc-fraction comigrated upon agarose gel electrophoresis with the extrachromosomal DNA band found in the total DNA (Fig. 1*a*, lanes 3 and 4).

Comparison with supercoiled DNA markers indicated that the plasmid, which was designated pFZ1, had a size of approximately 10·5 kb. A physical map of pFZ1 containing 31 restriction sites was constructed (Fig. 2). Digestion of pFZ1 with *Kpn*I resulted in the formation of three DNA fragments of 2·1, 3·7 and 4·7 kb, fragments I,

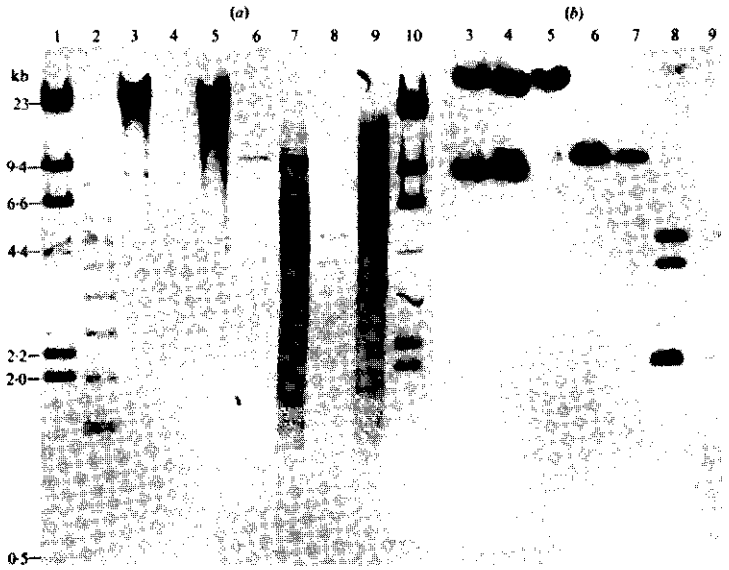


Fig. 1. Agarose gel electrophoresis (*a*) and Southern hybridization (*b*) of plasmid pFZ1 DNA purified by CsCl/ethidium bromide density-gradient centrifugation from *M. thermoformicicum* strain Z-245. Lanes 1 and 10, *Hin*dIII-digested phage λ DNA; lane 2, supercoiled size marker (with sizes from top to bottom of 16·2, 14·2, 12·1, 10·1, 8·1, 7·0, 6·0, 5·0, 4·0, 3·0 and 2·0 kb) (Life Technologies), lane 3, total undigested DNA of strain Z-245; lanes 4, 6 and 8, ccc plasmid DNA fraction, undigested, and digested by *Bam*HI and *Kpn*I, respectively; lanes 5, 7 and 9, oc fraction, undigested, and digested with *Bam*HI and *Kpn*I, respectively. The autoradiograph (*b*) is of the corresponding Southern blot hybridized under stringent conditions with equal amounts of *Kpn*I fragments I, II and III of plasmid pFZ1 labelled by nick-translation. The position of linear DNA size markers (in kb) is shown on the left.

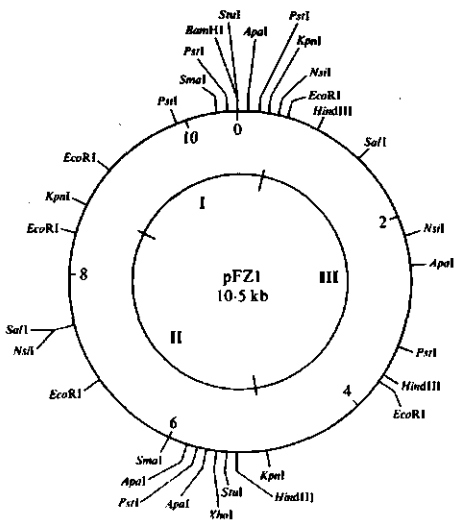


Fig. 2. Restriction map of plasmid pFZ1. The three subcloned *Kpn*I fragments I, II and III that were cloned using pUC19 are indicated in the inner circle. Sizes are given in kb.

II and III respectively (Fig. 2). All three *Kpn*I fragments were cloned in *E. coli* using pUC19 and used as probe in subsequent hybridization experiments. DNA purified from strain Z-245 by CsCl/ethidium bromide density-gradient centrifugation was digested with *Bam*HI and *Kpn*I, separated by agarose gel electrophoresis, blotted and hybridized to labelled pFZ1 fragments (Fig. 1*b*, lanes 6–9). Only the expected fragments of plasmid DNA present in the total DNA showed hybridization, indicating that no sequences homologous to plasmid pFZ1 are present in the chromosomal DNA of strain Z-245 (see also Fig. 3). Furthermore, using pFZ1 DNA as probe, no pFZ1 DNA could be detected in the supernatant of a culture of strain Z-245 with methods commonly used for the isolation of bacteriophage DNA (Maniatis *et al.*, 1982; Wood *et al.*, 1989).

We also detected plasmid DNA in *M. thermoformicicum* strains FTF and THF; these plasmids were designated pFZ2 and pFV1, respectively (Table 1). Plasmid pFZ2 was similar in size to plasmid pFZ1 (10·5 kb, see below). In contrast, plasmid pFV1 was larger than plasmid pFZ1, with a size of approximately 14 kb as revealed by comparison of ccc-DNA from plasmid pFV1 isolated by CsCl centrifugation with supercoiled DNA

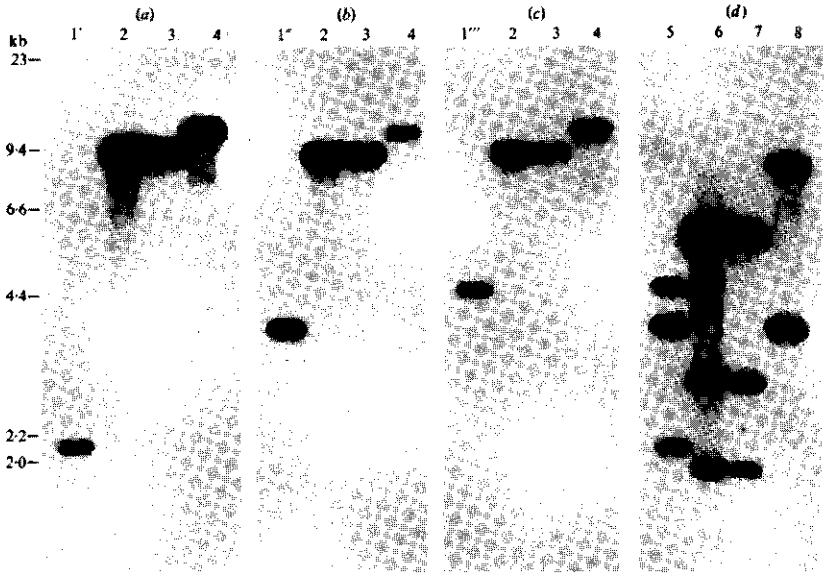


Fig. 3. Southern hybridization of total DNA from *M. thermoformicicum* strains Z-245, FTF and THF with labelled pFZ1 DNA. Total DNA (about 1·5 μg each) of strains Z-245 (lanes 2 and 6), FTF (lanes 3 and 7) and THF (lanes 4 and 8) was digested with *Xho*I (panels *a*, *b* and *c*) and *Hin*dIII (panel *d*) and separated by agarose gel electrophoresis. Blots were hybridized under stringent conditions to equal amounts of *Kpn*I fragments I (*a*), II (*b*) and III (*c*), and to all three fragments (*d*), labelled by nick-translation. Lanes 1', 1" and 1''' contain purified *Kpn*I fragments I, II and III of plasmid pFZ1, respectively, used as the probe; lane 5 contains all three *Kpn*I fragments. The position of linear DNA size markers (in kb) is shown on the left.

markers separated by agarose gel electrophoresis (data not shown). No extrachromosomal DNA was detected in *M. thermoformicicum* strains CB12 and SF-4 or in the new isolates, strains FF1 and FF3.

Determination of the copy-number of plasmid pFZ1

To determine the copy-number of plasmid pFZ1 several dilutions of total DNA isolated from strain Z-245 were analysed by agarose gel electrophoresis. By comparing the intensity of plasmid and chromosomal DNA bands we found that the intensity of the ccc-plasmid band in the undiluted fraction was comparable with that of the chromosomal DNA band in the fraction diluted 40 times. Assuming that the size of the chromosomal DNA of *M. thermoformicicum* strain Z-245 is equivalent to that of *M. thermoautotrophicum* strain Marburg ($1{\cdot}2 \times 10^9$ Da) (Brandis *et al.*, 1981), this indicates that strain Z-245 contains at least four copies of pFZ1 per chromosome.

Sequence homologies between plasmid pFZ1 and DNA from other methanogens

None of the labelled *Kpn*I-fragments of pFZ1 (Fig. 2) hybridized to total DNA of *M. thermoautotrophicum* strains Marburg and ΔH, *M. formicicum* strain MS1 or *M. thermoformicicum* strains CB12 and SF-4 and the new isolates FF1 and FF3 (data not shown). In contrast, hybridization of these pFZ1 DNA fragments was observed to total DNA isolated from strains FTF and THF. Subsequent experiments using undigested and digested DNA of both these strains showed that this was due to homology of pFZ1 DNA with the plasmid DNA of strains FTF and THF (data not shown).

To determine the degree of homology between the plasmids, total DNA isolated from *M. thermoformicicum* strains Z-245, FTF and THF was digested with *Xho*I and *Hin*dIII, separated by agarose gel electrophoresis, blotted and hybridized to the individually labelled *Kpn*I fragments I, II and III of pFZ1 (Fig. 2). The results (Fig. 3) indicate that (i) the hybridizing fragments of DNA

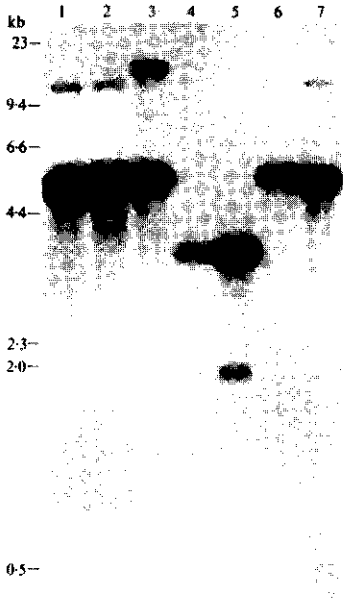


Fig. 4. Southern hybridization of total DNA of *M. thermoformicicum* with the *M. formicicum fdhA* gene. Total DNA (about 1·5 μg each) of strains Z-245 (lane 1), FTF (lane 2), THF (lane 3), CB12 (lane 4), SF-4 (lane 5), FF1 (lane 6) and FF3 (lane 7) was digested with *Bam*HI and separated by agarose gel electrophoresis. After transfer to a nylon membrane, the DNA was hybridized under moderately stringent conditions with a 2·2 kb *Sph*I–*Bam*HI fragment containing the *fdhA* gene from *M. formicicum* strain JF-1, labelled by nick-translation (Shuber *et al.*, 1986; kindly provided by J. G. Ferry, Department of Anaerobic Microbiology, VPI, Blacksburg, Va., USA). The position of linear DNA size markers (in kb) is shown on the left.

from strains Z-245 and FTF are of identical size (Fig. 3, lanes 2, 3, 6 and 7), corresponding with that expected from pFZ1 (Fig. 2); (ii) a different hybridization pattern is obtained with DNA from strain THF (Fig. 3, lanes 4 and 8); and (iii) that pFZ1 has no detectable homology with the chromosomal DNA of all three strains.

To analyse the high similarity of plasmid DNA from strains *M. thermoformicicum* Z-245 (pFZ1) and FTF (pFZ2) in more detail, total DNA from both strains was digested with restriction enzymes known to have several recognition sites in plasmid pFZ1 and, after separation and blotting, the DNA was hybridized to labelled pFZ1 DNA. All 31 restriction fragments derived from digestion of plasmid DNA of both strains with *Eco*RI, *Pst*I, *Apa*I, *Rsa*I, *Alu*I, *Hae*III and *Msp*I were found to have the same size (data not shown).

Hybridization of M. thermoformicicum strains to fdhA

As described above *M. thermoformicicum* strains Z-245, FTF and THF harbour plasmids that are (partially) homologous whereas other strains lack plasmid DNA. Using the *M. formicicum fdhA* gene as a hybridization probe, the relationship between the *M. thermoformicicum* strains was examined. As shown in Fig. 4 *M. thermoformicicum* strains fell into two groups on the basis of hybridization patterns of total DNA: one group, designated the Z-245 group, comprised strains Z-245, FTF and THF (lanes 1, 2 and 3); a second, designated the CB12 group, included strains CB12 and SF-4 (lanes 4 and 5). Within these two groups the hybridization patterns were identical except that strain THF showed a different minor hybridization species (Fig. 4, lane 3).

The *fdhA* gene was also hybridized to DNA of the new strains FF1 and FF3. These strains were assigned to *M. thermoformicicum* because of their capacity to produce methane at 55 °C using either H_2/CO_2 or formate as substrate, and their similar morphology to known *M. thermoformicicum* strains. The hybridization results (Fig. 4) show that strains FF1 and FF3 seem to be different from each other since the major hybridization species has a slightly different mobility in each case (lanes 6 and 7). Both strains, however, show very similar hybridization patterns that resemble that of strain Z-245, and therefore these new strains may be assigned to the Z-245 group (lanes 2, 6 and 7). Interestingly, both strains FF1 and FF3 lack plasmid DNA, in contrast to all other strains of the Z-245 group (Table 1).

Discussion

The present report describes the identification, isolation and characterization of extrachromosomal DNA from the thermophilic, formate-utilizing *M. thermoformicicum* strains Z-245, FTF and THF. The extrachromosomal DNA of strain Z-245 has the characteristics of plasmid DNA, being isolated in a ccc-form by CsCl/ethidium bromide density-gradient centrifugation and not detectable in the culture supernatant as described for virus-like particles of *Methanococcus voltae* (Wood *et al.*, 1989). In addition, pFZ1 shows no homology with the chromosomal DNA of the host strain Z-245 nor with that from other species of the genus *Methanobacterium* tested in this study (Figs 1 and 3). This is, to our knowledge, the first report of the presence of plasmid DNA in formate-utilizing, thermophilic methanogens. Plasmid pFZ1 is a relatively low-copy-number plasmid with a minimal four copies per chromosome of strain Z-245 and is stably maintained in its host during several serial transfers in the laboratory. At present we have not been able to

assign a specific function to pFZ1, which therefore remains cryptic, as are all other methanogen plasmids described so far.

Hybridization experiments revealed that *M. thermoformicicum* strains Z-245 and FTF harbour the very similar, if not identical, 10·5 kb plasmids pFZ1 and pFZ2, respectively. Both plasmids are homologous to plasmid pFV1 present in strain THF, which is clearly larger and has a different restriction fragment pattern. These results indicate that the plasmids of *M. thermoformicicum* share a common ancestor. The presence of highly similar plasmid DNA has been described for *Methanosarcina acetivorans* and two other marine acetotrophic strains, all three of which were isolated from the same location (Sowers & Gunsalus, 1988). In contrast, strains Z-245, FTF and THF were isolated from geographically very different locations (Touzel *et al.*, 1988; Zhilina *et al.*, 1984; Zinder & Koch, 1984).

Several attempts have been described to determine the relationship among *M. thermoformicicum* strains. Yamamoto *et al.* (1989) suggested that the strains may be divided into two groups, based on the cross-reaction of antibodies raised against strain SF-4 with strain CB12 and Z-245 but not with strain FTF. In contrast, DNA–DNA homology studies (J. P. Touzel, personal communication), indicated that strain FTF is more related to strains Z-245 and THF than to strain CB12 (strain SF-4 was not tested). Our observation that strains Z-245, FTF and THF harbour (partially) homologous plasmid DNA and the results based on the hybridization patterns with the *fdhA* gene support the classification of strains Z-245, FTF and THF into one group and strains CB12 and SF-4 into another group. On the basis of their hybridization patterns with the *fdhA* gene the new isolates FF1 and FF3 appear to be related to the Z-245 group. Since they are plasmid-free, those strains could be good candidates as recipients for plasmid pFZ1 or derivatives in transformation studies.

We wish to thank J. P. Touzel for providing strains that were used in an initial stage of the work described here. In addition, we are grateful to K. Yamamoto and G. D. Vogels for providing strains, J. G. Ferry for providing us with the *fdh* genes and R. Eggen for critical reading of the manuscript.

References

Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R. & Wolfe, R. S. (1979). Methanogens: re-evaluation of a unique biological group. *Microbiological Reviews* **43**, 260–296.

Brandis, A., Thauer, R. K. & Stetter, K. O. (1981). Relatedness of strain ΔH and Marburg of *Methanobacterium thermoautotrophicum*. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene* (Abteilung 1) **C2**, 311–317.

Brown, J. W., Daniels, C. J. & Reeve, J. N. (1989). Gene structure, organization, and expression in archaebacteria. *CRC Critical Reviews in Microbiology* **16**, 287–337.

DasSarma, S., Damerval, T., Jones, J. G. & Tandeau de Marsac, N. (1987). A plasmid-encoded gas vesicle protein in a halophilic archaebacterium. *Molecular Microbiology* **1**, 365–370.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Meile, L., Kiener, A. & Leisinger, T. (1983). A plasmid in the archaebacterium *Methanobacterium thermoautotrophicum*. *Molecular and General Genetics* **191**, 480–484.

Mountfort, D. O., Mörschel, E., Beimborn, D. B. & Schönheit, P. (1986). Methanogenesis and ATP synthesis in a protoplast system of *Methanobacterium thermoautotrophicum*. *Journal of Bacteriology* **168**, 892–900.

Pfeifer, F., Weidinger, G. & Goebel, W. (1981). Genetic variability of *Halobacterium halobium*. *Journal of Bacteriology* **145**, 375–381.

Schönheit, P., Moll, J. & Thauer, R. K. (1980). Growth parameters (K_s, μmas, Y_s) of *Methanobacterium thermoautotrophicum* strain Marburg. *Archives of Microbiology* **127**, 59–65.

Shuber, A. P., Orr, E. C., Recny, M. A., Schendel, P. F., May, H. D., Schauer, N. L. & Ferry, J. G. (1986). Cloning, expression, and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum*. *Journal of Biological Chemistry* **261**, 12942–12947.

Sowers, K. R. & Gunsalus, R. P. (1988). Plasmid DNA from the acetotrophic methanogen *Methanosarcina acetivorans*. *Journal of Bacteriology* **170**, 4979–4982.

Thomm, M., Altenbuchner, J. & Stetter, K. O. (1983). Evidence for a plasmid in a methanogenic bacterium. *Journal of Bacteriology* **153**, 1060–1062.

Touzel, J. P. & Albagnac, G. (1983). Isolation and characterization of *Methanococcus mazei* strain MC_3. *FEMS Microbiology Letters* **16**, 241–245.

Touzel, J. P., Petroff, D., Maestrojuan, G. M., Prensier, G. & Albagnac, C. (1988). Isolation and characterization of a thermophilic *Methanobacterium* able to use formate, the strain FTF. *Archives of Microbiology* **149**, 291–296.

Van Etten, J. C. & Freer, S. N. (1978). Simple procedure for disruption of fungal spores. *Applied and Environmental Microbiology* **35**, 622–623.

Wilharm, T., Thomm, M. & Stetter, K. O. (1986). Genetic analysis of plasmid pMP1 from *Methanolobus vulcani*. *Systematic and Applied Microbiology* **7**, 401.

Winter, J., Lerp, C., Zabel, H. B., Wildenhauer, F. X., König, H. & Schneider, F. (1984). *Methanobacterium wolfei*, sp. nov., a new tungsten requiring thermophilic, autotrophic methanogen. *Systematic and Applied Microbiology* **5**, 457–466.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archae, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 4576–4579.

Wood, A. G., Whitman, W. B. & Konisky, J. (1985). A newly-isolated marine methanogen harbors a small cryptic plasmid. *Archives of Microbiology* **142**, 259–261.

Wood, A. G., Whitman, W. B. & Konisky, J. (1989). Isolation and characterization of an archaebacterial virus like particle from *Methanococcus voltae* A3. *Journal of Bacteriology* **171**, 93–98.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Yamamoto, K., Tachibana, A., Dhavises, G., Tanaka, T., Tanigu-chi, M. & Oi, S. (1989). Characterization of a thermophilic formate-utilizing methanogen, *Methanobacterium thermoformicicum* strain SF-4. *Agricultural and Biological Chemistry* **53**, 533–543.

Zhao, H., Wood, A. G., Widdel, F. & Bryant, M. P. (1988). An extremely thermophilic *Methanococcus* from a deep sea hydrothermal vent and its plasmid. *Archives of Microbiology* **150**, 178–183.

Zhilina, T. N. & Ilaridnov, S. A. (1984). Characteristics of formate-assimilating methane bacteria and description of *Methanobacterium thermoformicicum* sp. nov. *Microbiology* **53**, 647–651 (English translation of *Mikrobiologyia* **53**, 785–790.

Zinder, S. H. & Koch, M. (1984). Non-acetoclastic methanogenesis from acetate: acetate oxidation by a thermophilic syntrophic coculture. *Archives of Microbiology* **138**, 263–272.

Modular organization of related archaeal plasmids encoding different restriction-modification systems in *Methanobacterium thermoformicicum*

Jörk Nölling, Freek J. M. van Eeden, Rik I. L. Eggen and Willem M. de Vos

Nucleic Acids Research **20**, 6501-6507.

Nucleic Acids Research, 1992, Vol. 20, No. 24 6501–6507

Modular organization of related Archaeal plasmids encoding different restriction – modification systems in *Methanobacterium thermoformicicum*

Jörk Nölling, Freek J.M.van Eeden, Rik I.L.Eggen and Willem M.de Vos*
Bacterial Genetics Group, Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

Received October 8, 1992; Revised and Accepted November 23, 1992 EMBL accession nos X68366, X68367

ABSTRACT

Nucleotide sequence comparison of the related 13513-bp plasmid pFV1 and the 11014-bp plasmid pFZ1 from the thermophilic archaeon *Methanobacterium thermoformicicum* THF and Z-245, respectively, revealed a homologous, approximately 8.2 kb backbone structure that is interrupted by plasmid-specific elements. Various highly conserved palindromic structures and an ORF that could code for a NTP-binding protein were identified within the backbone structure and may be involved in plasmid maintenance and replication. Each plasmid contains at comparable locations a module which specifies components of different restriction – modification (R/M) systems. The R/M module of pFV1 contained, in addition to the genes of the GGCC-recognizing R/M system *Mth*TI, an ORF which may be involved in repair of G-T mismatches generated by deamination of m^5C at high temperatures.

INTRODUCTION

Complete nucleotide sequence analysis of genomes may provide insight into their organization and evolution. In contrast to their bacterial counterparts, only limited information exists on extrachromosomal elements from members of the domain *Archaea*, a group of prokaryotic organisms that differs fundamentally from *Bacteria* (1, 2). Nucleotide sequences of only a few extrachromosomal elements isolated from these organisms have been analyzed, including the 15.5-kb genome of the *Sulfolobus shibatae* virus SSV1 (3), three small plasmids, pHSB1 (1.7 kb), pHGN1 (1.8 kb) and pGRB1 (1.8 kb), isolated from extreme halophilic bacteria (4, 5, 6), and the 4.4-kb pME2001 from the methanogen *Methanobacterium thermoautotrophicum* strain Marburg (7).

Recently, a family of three related plasmids was identified in the thermophilic archaeon *Methanobacterium thermoformicicum* (8) comprising the nearly identical 11-kb pFZ1 and pFZ2, harbored by strains Z-245 and FTF, respectively, and the closely related 13.5 kb pFV1 of strain THF. In addition to the high interplasmid similarity, certain plasmids regions were identified with homology to chromosomal DNA from different thermophilic *Methanobacterium* strains that lack plasmid DNA (9). The plasmids may confer a selective advantage to their hosts since it has been shown that restriction and modification (R/M) systems were encoded by pFV1 (10) and pFZ1 and pFZ2 (11).

This report describes the complete nucleotide sequences of pFV1 and pFZ1 and their comparison. Several highly conserved regions were identified that may be involved in replication and maintenance of either plasmid. In addition, a putative DNA-repair enzyme encoded by pFV1 was identified.

MATERIAL AND METHODS

Bacterial strains, plasmids and growth conditions

M.thermoformicicum strains THF and Z-245 were cultivated as described previously on H_2/CO_2 as sole carbon and energy source (8). *Escherichia coli* strains DH5α (Life Technologies Inc., Gaithersburg, Md.) and TG1 (12) were used for propagation of plasmids pUC18 or pUC19 and phages M13mp18 or M13mp19 (13), respectively, and strain Q359 (14) served as host for propagation of bacteriophage λ GEM-12 (Promega Biotec, Madison, Wis.). All *E.coli* strains were grown at 37°C in Luria broth-based media (15). If appropriate, ampicillin (50 μg/ml), isopropyl-β-D-thiogalactopyranoside (1 mM), or 5-bromo-4-chloro-3-indolyl-β-D-galactoside (0.004%) was added.

Chemicals and enzymes

[α-^{32}P]dATP (110 TBq/mmol) was obtained from Amersham International (Amersham, United Kingdom). All enzymes used for manipulation of nucleic acids were purchased from Life Technologies, Inc. or Pharmacia LKB Biotechnologies (Uppsala, Sweden). Oligonucleotides were synthesized with a Biosearch Cyclone DNA synthesizer (New Brunswick Scientific Corp.) at the Netherlands Institute for Dairy Research (NIZO) or were purchased from Pharmacia (Uppsala, Sweden). Other chemicals were obtained in PA quality from Sigma Chemical Co. (St. Louis, Mo.), from Boehringer GmbH (Mannheim, FRG) or from Merck (Darmstadt, FRG).

* To whom correspondence should be addressed

DNA isolation and PCR amplification

Total DNA from methanogenic bacteria was isolated as described previously (8). A pFV1 fragment that spans the single *Xho*I site (position 0/13513) was amplified by PCR as described earlier (16) using total DNA from *M.thermoformicicum* strain THF and two oligonucleotides, 5'-CTGTAGTTCAGGATC-3' and 5'-GAGATACTGTTCAACTG-3', complementary to positions 377 to 363 and 12921 to 12937 of plasmid pFV1, respectively. The amplification product was treated with Klenow DNA polymerase for 30 min at room temperature in the presence of all four dNTPs (2 mM each), followed by an incubation for 20 min at 55°C in the presence of proteinase K (0.1 mg/ml) and sodium dodecyl sulfate (0.5%) essentially as described by Crowe and coworkers (17). Subsequently, the suspension was extracted with phenol/chloroform, the DNA was precipitated with ethanol, and dissolved in TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). After digestion of the amplified DNA with *Hind*II and *Pst*I, the 0.4-kb fragment that contained the *Xho*I-site was recovered from agarose gels using the Geneclean kit (Bio 101, La Jolla, Calif.) and cloned into M13mp19.

DNA manipulations and sequence analysis

DNA manipulations were performed essentially as described by Sambrook and coworkers (15). For the determination of the nucleotide sequence, fragments of cloned pFZ1 (8) and pFV1 DNA (see below) were isolated from agarose gels and subcloned into M13mp18/19 vectors. Sequence reactions with vector- or insert specific primers were carried out by the dideoxy chain termination method (18) using either reagents from the Sequenase kit (United States Biochemical Corp., Cleveland, Ohio) according to the recommendations of the manufacturer or reagents for Taq-polymerase sequencing (19). The nucleotide sequence of plasmid pFV1 was determined on both strands apart from two small regions (positions 876–1182 and 2392–2786) that could not be sequenced in both directions due to a premature termination under all conditions. For plasmid pFZ1 the nucleotide sequence of overlapping fragments was determined on one strand and for 40% on both strands. Computer analysis of sequence data was done with the University of Wisconsin Genetics Computer Group package version 6.0 (20) and the CAOS/CAMM facilities at Nijmegen, The Netherlands.

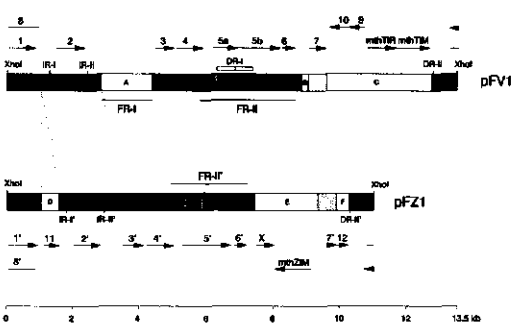


Figure 1. Comparison of the genetic organization of *Xho*I-linearized pFV1 and pFZ1 from *M.thermoformicicum* THF and Z-245, respectively. The degree of interplasmid homology is indicated by shading of the maps: the darkly shaded regions show high sequence similarity whereas the white regions A through F represent plasmid-specific sequences. Each ORF listed in Table 1 is shown by a line with an arrowhead indicating its direction. The positions of regions that contain direct (DR) or inverted (IR) repeats are indicated. Dotted lines show positions of fragment insertions. FR-I and FR-II represent regions with homology to chromosomal DNA from several thermophilic strains of the genus *Methanobacterium*.

RESULTS AND DISCUSSION

Cloning of plasmid pFV1

In a previous report we have shown that *M.thermoformicicum* strain THF harbors the low copy number plasmid pFV1 (8). To be able to analyze this plasmid in detail, covalently closed circular plasmid pFV1 DNA (8) was linearized with *Xho*I and ligated to *Xho*I-digested λ GEM-12. The ligation mix was packaged *in vitro* (15) and used to infect *E.coli* Q359. A 2.3-kb *Kpn*I-fragment of plasmid pFZ1 that strongly hybridized to plasmid pFV1 (8), was labelled by nick-translation and used as probe to screen the recombinant phages. A strongly hybridizing recombinant phage, designated λ GFV1, was identified and, after plaque purification, phage DNA was prepared (15). The integrity of the pFV1 insert of λ GFV1 was verified by restriction analysis and hybridization experiments (data not shown). Purified λ GFV1 DNA was used as source of pFV1 DNA in subsequent experiments.

Nucleotide sequence of plasmid pFV1 and pFZ1 and their comparison

The complete nucleotide sequences of the plasmids pFV1 (EMBL accession number X68366) and pFZ1 (EMBL accession number X68367) from *M.thermoformicicum* strains THF and Z-245, respectively, were determined. The single *Xho*I-site of each plasmids was taken as position 1. Since plasmid pFV1 was originally cloned as a *Xho*I fragment (see above), the uniqueness of this restriction site was confirmed by sequence analysis of a PCR amplified DNA fragment of pFV1 that overlaps the *Xho*I-site. Plasmid pFV1 had a size of 13513 bp and a calculated GC content of 41.8% which is significantly lower than that of its host strain THF (49.6%) (21).

The pFZ1 genome was 11014 bp in length. Alignment of the nucleotide sequences of plasmid pFZ1 and pFV1, schematically shown in Figure 1, revealed a high degree of similarity, confirming previous hybridization experiments (8), and a comparable sequence organization.

Several regions of considerable similarity were identified which are interrupted by plasmid-specific sequences. The homologous sequence blocks, the majority of which shares more than 91% identical nucleotides, add up to approximately 8.2 kb and constitute the backbone structure of both plasmids (Figure 1). As a consequence, these regions have comparable coding capacities and may specify proteins with more than 84% identical amino acids (Table 1).

Two plasmid regions, designated FR-I and FR-II (Figure 1), were found to have homologous counterparts in the chromosomal DNA of several thermophilic *M.thermoformicicum* strains that lack plasmid DNA (9). FR-I resembles an insertion sequence (9) and is specific for plasmid pFV1 (see below), whereas FR-II is part of both plasmids. FR-II of pFV1, however, contains a direct repeat of two identical 524-bp sequences (DR-I, position 6374 to 7421; Figure 1) that is not found in the corresponding region of plasmid pFZ1.

The most conspicuous difference between the plasmids pFV1 and pFZ1 is the presence of three unique sequences that could

be identified in each plasmid (element A–F in Figure 1). The location of two of those elements, A and D, in regions with high interplasmid similarity allowed to determine their exact size. The AT-rich element A (GC content 29.7%), which corresponds to FR-I (Figure 1), is located on plasmid pFV1 (position 2852 to 4352) and comprises 1501 bp. Element D is a 569-bp insertion of plasmid pFZ1 (position 917 to 1485; Figure 1) and contains open reading frame (ORF) 11 (Figure 1). Four other unique sequences, B and C of plasmid pFV1 and E and F of plasmid pFZ1, are located at similar positions relative to the plasmid backbone and flank a region of approximately 0.5 kb that exhibits only relatively low interplasmid similarity (69%; Figure 1). The two large plasmid-specific elements C and E of pFV1 and pFZ1, respectively, code for components of recently characterized R/M systems: the GGCC-recognizing *Mth*TI system of plasmid pFV1 comprising the methyltransferase (MTase) gene *mthTIM* and the endonuclease (ENase) gene *mthTIR* (10), and the CTAG-recognizing system *Mth*ZI of plasmid pFZ1 encoded by the MTase gene *mthZIM* and ORF X which may code for the corresponding ENase (11) (Figure 1). Moreover, two additional ORFs, ORF9 and 10, are located on the unique pFV1-fragment C in close proximity to the genes of the *Mth*TI system, one of which, ORF10, may encode a protein with considerable similarity to enzymes involved in DNA mismatch repair (see below).

Repeated sequences of plasmid pFV1 and pFZ1

The close relationship of pFV1 and pFZ1 (Figure 1) implies a similar replication mechanism for both replicons that should be specified by the conserved backbone structures. Since repeated sequences are known to be involved in the replication of bacterial plasmids (for reviews: 22, 23), we searched in the nucleotide sequence of pFV1 and pFZ1 for regions that contain direct (DR) or inverted repeats (IR).

Apart from DR-I of plasmid pFV1 described above, a region with several DRs was identified in the backbone structure of both plasmids (Figures 1). This 490-bp segment of plasmid pFV1 (DR-II; position 12681 to 13170) contains three perfect 51-bp and two nearly perfect 33-bp repeats. Comparison with the corresponding region of plasmid pFZ1 (DR-II', position 10265 to 10671) showed, however, only limited similarity, and only the two 33-bp repeats were found to be almost completely conserved.

Inspection of the pFV1 and pFZ1 nucleotide sequences revealed two IRs, IR-I and IR-II, that are contained within highly conserved, non-coding regions of both plasmids (Figure 1). Both IR-I and IR-II exhibit a high GC-content of 52% GC and 62% GC, respectively, and have the capacity to form secondary structures. In IR-I (position 1102 to 1419 in pFV1; Figures 1 and 2), four potential hairpin structures (IR1-IR4, Figure 2) could be identified. Remarkably, the hairpin loops formed by IR1 and IR4 contain a partially overlapping pentamer with the sequence 5'-TCGCT which is repeated four times in the loop of IR1 and five times in the loop of IR4 (Figure 2).

The second region with palindromic sequences, IR-II, is located approximately 1 kb downstream of IR-I (ranging from position 2392 to 2593 in plasmid pFV1) and contains two palindromic sequences, IR5 and IR6 (Figure 2). Analogous to IR-I, two partially overlapping repeats of the pentamer TCGCT are present in both the hairpin loop and the stem/interior loop of IR6, whereas perfect repeats of this pentamer are also found in the region between IR5 and IR6 (Figures 1 and 2).

Both IR-I and IR-II represent one of the most conserved plasmid regions in pFV1 and pFZ1 that differ at only 11 positions out of 520 nucleotides (Figure 2). This high conservation of the IR-regions may indicate an essential function in plasmid maintenance or replication functions. Interestingly, in plasmids that replicate by a rolling circle mechanism via a single-stranded intermediate, (imperfect) palindromic structures are part of the so-called plus and minus origin, the former of which is used for initiation of replication by a plasmid-encoded Rep protein (22). Moreover, the hairpin loops of the minus origins were found to share the consensus sequence 5'-TAGCGT (24). This sequence resembles the complementary sequence of the repeated pentamer 5'-TCGCT which has been observed in hairpin loops of IR-I and IR-II of both plasmids pFV1 and pFZ1. However, we neither could detect single-stranded pFV1- or pFZ1 DNA in hybridization experiments (data not shown) nor found similarity of any of the plasmid-backbone encoded proteins with conserved sequence motifs of rolling circle replication initiator proteins (25). Therefore, the involvement of the IR-regions in replication remains to be confirmed.

Coding capacity and putative genes of plasmid pFV1 and pFZ1

Analysis of the nucleotide sequences of plasmid pFV1 and pFZ1 for coding regions revealed 13 and 12 ORFs, respectively, that could code for polypeptides composed of more than 100 amino

Table 1. Comparison of (putative) genes located on plasmids pFV1 and pFZ1

Gene	plasmid pFV1			Gene	plasmid pFZ1			% Identity
	Amino acids	Start codon	DNA sequence		Amino acids	Start codon	DNA sequence	
ORF1	364	ATG	13304–882	ORF1'	364	ATG	10805–882	99
ORF2	284	ATG	1511–2362	ORF2'	284	ATG	2101–2952	85
ORF3	197	ATG	4438–5028	ORF3'	217	ATG	3455–4105	98
ORF4	237	ATG	5152–5862	ORF4'	234	ATG	4225–4926	91
ORF5a	227	ATG	6229–6909	ORF5'	499	ATG	5295–6791	72
ORF5b	451	GTG	6912–8264	ORF5'	499	ATG	5295–6791	84
ORF6	122	ATG	8354–8719	ORF6'	122	ATG	6894–7259	98
ORF7	145	ATG	9186–9620	ORF7'	80	ATG	9647–9886	71
ORF8	361	ATG	822–13253	ORF8'	361	ATG	822–10754	88
ORF9	146	ATG	10733–10296	ORF11	148	ATG	1093–1536	–
ORF10	221	TTG	10300–9638	ORF12	117	GTG	9889–10239	–
mthTIR	281	ATG	10827–11669	ORFX	202	GTG	7539–8144	–
mthTIM	330	ATG	11669–12658	*mthZIM*	355	ATG	9203–8139	–

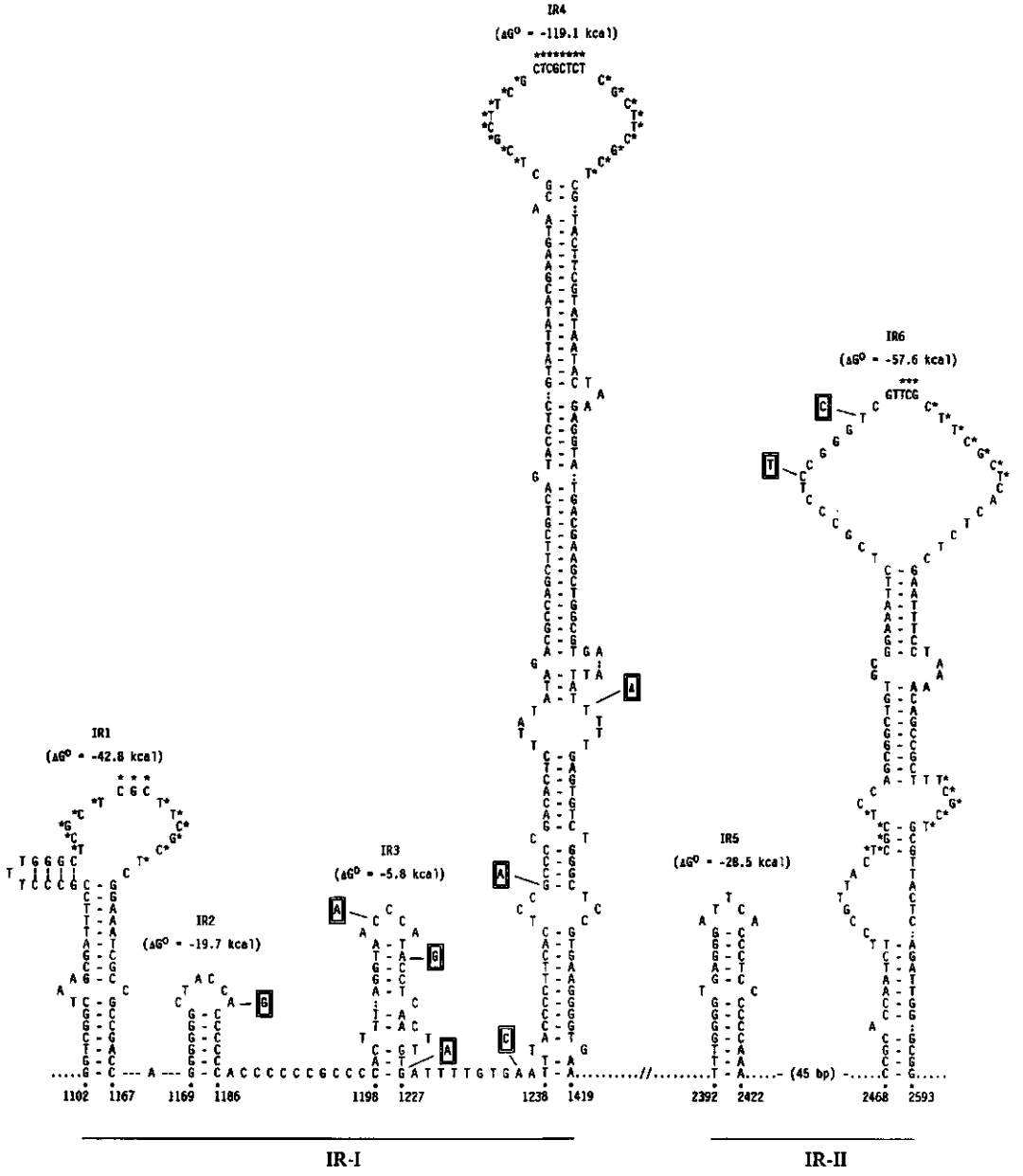


Figure 2. Potential secondary structures of regions IR-I and IR-II and their locations in pFV1. The free energy (ΔG°) values of the potential hairpin loops IR1 to IR6 were calculated using the PCGENE program RNAFOLD. Nucleotides different with the corresponding regions of pFZ1 are indicated. The repeated pentamers 5'-TCGCT are marked by asterisks.

acids (Figure 1; Table 1). Except for ORF8 of pFV1 and the homologous ORF8′ of pFZ1, all ORFs located in the backbone structure have the same orientation. Each plasmid specifies components of different R/M systems (10, 11) which are encoded by genes located on the plasmid specific elements C (pFV1) and E (pFZ1) (Figure 1). The deduced amino acid sequences of the other ORFs of pFV1 and pFZ1, listed in Table 1, were analyzed for similarity with sequences present in the SwissProt data base (release 22.0).

The homology search showed that ORF1 of plasmid pFV1 and its counterpart ORF1′ of plasmid pFZ1 contain a stretch of amino acids with similarity to the purine nucleotide binding site of ATP- or GTP-utilizing proteins (Table 2). This NTP-binding site is formed by a characteristic amino acid motif (NTP-motif) composed of two elements A and B (26, 27) and—analogous to several NTP-binding enzymes involved in DNA-metabolism (27–30)—is located within the N-terminal part of the ORF1 and ORF1′ proteins (Table 2). Together with the high conservation of the deduced ORF1 and ORF1′ proteins (>99% identical residues; Table 1), the presence and location of this NTP-binding motif strongly suggests that both proteins are involved in the replication of pFV1 and pFZ1. Transcription and translation of ORF1 and ORF1′ may be initiated at sequence motifs that were found preceding the coding regions. A sequence stretch with similarity to a typical archaeal promoter signal (31, 32) was identified ranging from position 13228 to 13258 in pFV1 and from position 10729 to 10759 in pFZ1. The first translation initiation codon (ATG) is preceded by a sequence with similarity to a consensus methanogenic ribosome binding site (33), ranging from position 13294 to 13300 in pFV1 and from position 10795 to 10801 in pFZ1. However, no sequences resembling the consensus archaeal terminator signal (33) have been recognized downstream of ORF1 and ORF1′.

Within the pFV1-specific DNA element C which contains the genes for the *Mth*TI R/M system, two other ORFs, designated ORF9 and ORF10, were identified that could code for proteins of 146 and 221 amino acids, respectively (Figure 1; Table 1). A database search revealed that the protein deduced from ORF10 shares significant sequence similarity with two *E.coli* proteins, MutY and endonuclease III (52.4% and 50.0% similarity and 28.8% and 28.4% identity), respectively, which are known to be involved in DNA mismatch repair. MutY and endonuclease III, composed of 350 (34) and 211 amino acids (35), respectively, both possess N-glycosylase activity as well as apurinic/apyrimidinic (AP) endonuclease activity (36–39). They differ, however, in their substrate specificity. While MutY is involved in the correction of A/G (35), A/C (37) and A/8-oxoG

Table 2. Alignment of the NTP-motif containing domain of ORF1 with NTP-binding proteins involved in DNA metabolism. Identical or similar residues are boxed.

Protein		Sequence A		Sequence B	Reference
RecD	166	SV I S G GP G T GKT TTVARL L	-79-	V L V VDE ASMID	(28)
UvrD	24	LL V L A GA G S GKT RVLVKR I	-171-	N I L VDE FQMTN	(29)
RuvB	57	LL I F G PP G L GKT TLANIV A	-32-	V L F IDE IRRLS	(30)
ORF1	61	LL I V G PP G S GKT VTTKYV I	-54-	I I V HDE IDKTL	this study

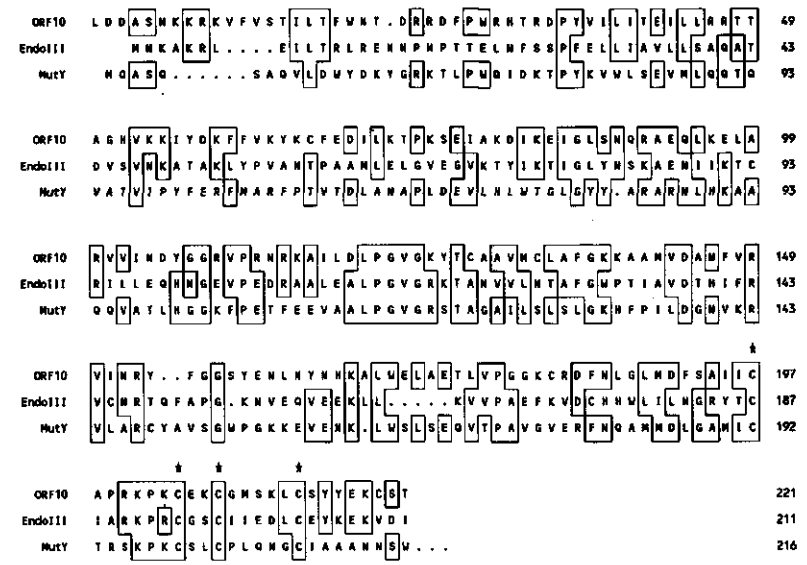


Figure 3. Amino acid sequence alignment of the putative ORF10 protein of pFV1 with endonuclease III and MutY of *E.coli*. Identical amino acids are boxed. Dots represent gaps that have been introduced to optimize the alignment. Positions marked with asterisks represent the four highly conserved cysteine residues. Numbers at the right refer to amino acid residues.

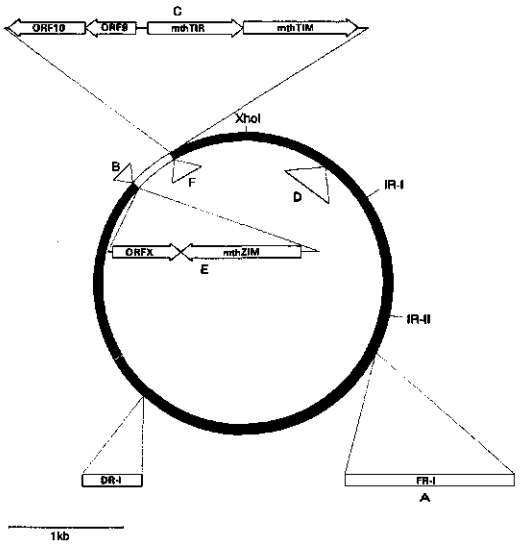


Figure 4. Schematic representation of the conserved 8.2 kb backbone structure obtained from nucleotide sequence alignment of pFV1 and pFZ1. Insertions indicated outside the circle represent pFV1-specific sequences, those inside are specific for pFZ1 (the bar represents a size of 1 kb).

mismatches (8-oxoG = 8-hydroxyguanine) (40), endonuclease III is active on mismatches generated by certain derivatives of thymidine and cytosine (38, 39). Alignment of the deduced amino acid sequences of the ORF10 protein, endonuclease III and MutY (Figure 3) revealed (i) a highly conserved motif LPGVG(R/K)XT (position 120 to 127 of the ORF10 protein; X = variable amino acid), and (ii) a stretch of four cysteine residues spaced by an identical number of amino acids ($CX_6CX_2CX_5C$), which is located at corresponding positions in the three polypeptides. Since endonuclease III was shown to be an iron-sulfur protein (41), those cysteine residues are believed to be involved in the $[4Fe\text{-}4S]^{2+}$ cluster binding (41). It therefore seems to be likely that MutY, as already suggested by Michaels and coworkers (34), and the putative ORF10 protein specified by plasmid pFV1 are iron-sulfur proteins with $[4Fe\text{-}4S]^{2+}$ clusters. The observed sequence similarities suggest that the ORF10 protein acts, like MutY and endonuclease III, as a DNA repair enzyme that recognizes incorrect base pairing.

The presence of such a repair system in the thermophilic *M.thermoformicicum* strain THF is plausible since the pFV1-encoded GGCC-recognizing methyltransferase M.*Mth*TI generates 5-methylcytosine (m^5C) (10) which is, especially under thermophilic conditions, subject to deamination (m^5C T) resulting in a G-T mismatch (42). To avoid the probably lethal effect of multiple transitions, strain THF should contain a mechanism that either prevents deamination of m^5C or is capable to correct the G-T mismatches. The pFV1-encoded ORF10 protein is a good candidate to act as a G-T mismatch repair enzyme in *M.thermoformicicum* THF. Moreover, the close proximity of ORF10 and the *Mth*TI system would support a functional relation of the endonuclease, the methyltransferase and the putative DNA mismatch repair enzyme. Such a clustering may also facilitate the simultaneous transfer of these genes between cells and ensures the coordinate activity of the encoded enzymes in the recipient which probably is necessary for its survival. Interestingly, a similar genetic linkage has been found for the functionally related methyltransferase and DNA mismatch repair enzyme in the *dcm-vsr* system of *E.coli* (42).

It seems paradoxical that *M.thermoformicicum* THF maintains a non-chromosomally encoded R/M system with a m^5C-producing MTase activity and, as a consequence, a DNA mismatch repair enzyme. A possible explanation is that, under thermophilic conditions, a combination of the enzymes specified by the GGCC-recognizing *Mth*TI R/M system and ORF10 may serve as a protection mechanism against infection by phages that exhibit the same methylation pattern as produced by the MTase M.*Mth*TI, i.e. $GG^{me}CC$. Such phages may have escaped the restriction barrier of strain THF or have acquired a protecting MTase gene. Since their methylated DNA is resistant against digestion by the host-encoded ENase, such phages normally would be able to infect the entire population. However, as a result of elevated deamination of m^5C at high temperatures without mispair correction, the genome of those phage particles accumulate, in contrast to the host genome, partially and fully deaminated $GG^{me}CC$-sites the latter of which would result in two juxtaposed G-T mismatches per site. One may envisage several scenarios once this phage DNA has entered the host cell. Firstly, the ORF10 protein could be unable to initiate mispair correction at juxtaposed G-T mismatches which may decrease the infectivity of the damaged phage DNA. The second scenario is based on the observation that the *mut*HLS-dependent mismatch repair systems of certain *E.coli* mutants produce double-strand breaks due to initiation of mismatch correction along with an endonucleolytic attack on either DNA strand (44, 45). Double-strand breaks in the deaminated phage DNA may be generated during initiation of mismatch correction by the putative ORF10 protein which—analogous to MutY and endonuclease III—may act as N-glycosylase and AP-endonuclease on the opposite strands of the juxtaposed G-T mismatch base pairs. This would generate a double-stranded break 3′ from the GG-dinucleotide of the GGCC site resulting in a similar effect as restriction of the unmethylated phage DNA by the ENase of the R/M system *Mth*TI.

CONCLUSION

Sequence analysis of the plasmids pFV1 and pFZ1 from the archaeon *M.thermoformicicum* strain THF and Z-245, respectively, has confirmed their close relationship. Both plasmids display a modular genome organization and consist of a highly conserved, approximately 8.2 kb backbone which is interspersed with plasmid-specific, accessory elements of variable size and function (Figure 4).

The high sequence similarity and comparable genetic maps of the plasmids pFV1 and pFZ1 indicate either the presence of a common ancestor or that one of the plasmids originated from the other. In either case, accessory elements have been acquired or lost.

The conserved backbone structure is expected to specify function necessary for plasmid replication and maintenance. These essential plasmid functions may include two regions, IR-I and IR-II (Figures 2 and 4), with the capacity to form highly conserved secondary structures, and, in close proximity to those

IR-regions, ORF1 that could encode a protein with NTP-binding activity (Figure 1, Table 2).

While the nature of the smaller elements specific for pFV1 and pFZ1 remains unknown, element A of plasmid pFV1, which corresponds to FR-1, has various features of an insertion sequence (9). In addition, each plasmid contains, analogous to the functional modules observed for lambdoid phages (46), at comparable locations an element which specifies components of a R/M system recognizing either the sequence GGCC (pFV1) (10) or CTAG (pFZ1) (11). Within the R/M module of plasmid pFV1, two additional ORFs were identified, one of which, ORF10, may code for a DNA-repair enzyme that is functionally related to the GGCC-recognizing R/M genes.

REFERENCES

1. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA **287**, 4576–4579.
2. Zillig, W. (1991) Current Opinion in Genetics and Development **1**, 544–551.
3. Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D. and Zillig, W. (1991) Virology **185**, 242–250.
4. Hackett, N.R., and DasSarma, S. (1989) Can. J. Microbiol. **35**, 86–91.
5. Hall, M.J. and Hackett, N.R. (1989) Nucleic Acids Res. **17**, 10501.
6. Hackett, N.R., Krebs, M.P., DasSarma, S., Goebel, W., RajBhandary, U.L. and Khorana, H.G. (1990) Nucleic Acids Res. **18**, 3408.
7. Bokranz, M., Klein, A. and Meile, L. (1990) Nucleic Acids Res. **18**, 363.
8. Nölling, J., Frijlink, M. and de Vos, W.M. (1991) J. Gen. Microbiol. **137**, 1981–1986.
9. Nölling, J., van Eeden, F.J.M., and de Vos, W.M. submitted for publication.
10. Nölling, J. and de Vos, W.M. (1992) J. Bacteriol. **174**, 5719–5726.
11. Nölling, J. and de Vos, W.M. (1992) Nucleic Acids Res. **20**, 5047–5052.
12. Gibson, T.J. (1983) Ph.D. thesis. University of Cambridge. Cambridge, United Kingdom.
13. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
14. Kam, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Natl. Acad. Sci. USA **77**, 5172.
15. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Nölling, J., Hahn, D., Ludwig, W., and de Vos, W.M. submitted for publication.
17. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D. and Gewert, D. (1991) Nucleic Acids Res. **19**, 184.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
19. Innis, M., (1988) Proc. Natl. Acad. Sci. USA **85**, 9436–9440.
20. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395.
21. Touzel, J.P., Conway de Macario, E., Nölling, J., de Vos, W.M., Zhilina, T., and Lysenko, A.M. (1992) Int. J. Syst. Bacteriol. **42**, 408–411.
22. Gruss, A. and Ehrlich, S.D. (1989) Microbiol. Reviews **53**, 231–241.
23. Kües, U. and Stahl U. (1989) Microbiol. Reviews **53**, 491–516.
24. del Solar, G.H., Puyet, A. and Espinosa, M. (1987) Nucleic Acids Res. **15**, 5561–5580.
25. Ilyina, T. and Koonin, E.V. (1992) Nucleic Acids Res. **20**, 3279–3285.
26. Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) EMBO J. **1**, 945–951.
27. Hodgman, T.C. (1988) Nature (London) **333**, 22–23.
28. Finch, P.W., Storey, A., Brown, K., Hickson, I.D. and Emmerson, P.T. (1988) Nucleic Acids Res. **14**, 8583–8594.
29. Finch, P.W. and Emmerson, P.T. (1984) Nucleic Acids Res. **12**, 5789–5799.
30. Benson, F.E., Illing, G.T., Sharples, G.J. and Lloyd, R.G. (1988) Nucleic Acids Res. **16**, 1541–1549.
31. Hausner, W., Frey, G. and Thomm, M. (1991) J. Mol. Biol. **222**, 495–508.
32. Hain, J., Reiter, W.-D., Hüdepohl, U. and Zillig, W. (1992) Nucleic Acids Res. **20**, 5423–5428.
33. Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) Crit. Rev. Microbiol. **16**, 287–337.
34. Michaels, M.L., Pham, L., Nghiem, Y., Cruz, C. and Miller, J.H. (1990) Nucleic Acids Res. **18**, 3841–3845.
35. Asahara, H., Wistort, P.M., Bank, J.F., Bakerian, R.H. and Cunningham, R.P. (1989) Biochemistry **28**, 4444–4449.
36. Au, K.G., Clark, S., Miller, J.H., and Modrich, P. (1989) Proc. Natl. Acad. Sci. USA **86**, 8877–8881.
37. Tsai-Wu, J.-J., Liu, H.-F., and Lu, A.-L. (1992) Proc. Natl. Acad. Sci. USA **89**, 8779–8783.
38. Katcher, H.L. and Wallace, S.S. (1983) Biochemistry **22**, 4071–4081.
39. Kim, J. and Linn, S. (1988) Nucleic Acids Res. **16**, 1135–1141.
40. Michaels, M.L., Cruz, C., Grollman, A.P. and Miller, J.H. (1992) Proc. Natl. Acad. Sci. USA **89**, 7022–7025.
41. Cunningham, R.P., Asahara, H. and Bank, J.F. (1989) Biochemistry **28**, 4450–4455.
42. Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. (1985) Nucleic Acids Res. **13**, 1399–1412.
43. Sohail, A., Lieb, M., Dar, M. and Bhagwat, A.S. (1990) J. Bacteriol. **172**, 4214–4221.
44. Glickman, B.W. and Radman, M. (1980) Proc. Natl. Acad. Sci. USA **77**, 1063–1067.
45. Wang, T.V. and Smith, K.C. (1986) J. Bacteriol. **165**, 1023–1025.
46. Susskind, M.M. and Botstein, D. (1978) Microbiol. Rev. **42**, 385–413.

Distribution and characterization of plasmid-related sequences in the chromosomal DNA of different thermophilic *Methanobacterium* strains

Jörk Nölling, Freek J. M. van Eeden and Willem M. de Vos

SUMMARY

The genomes of several thermophilic members of the genus *Methanobacterium* were analyzed for homology to the related restriction-modification plasmids pFV1 and pFZ1 from *M.thermoformicicum* strains THF and Z-245, respectively. Two plasmid-located regions, designated FR-I and FR-II, could be identified with chromosomal counterparts in six *Methanobacterium* strains. Multiple copies of the pFV1-specific element FR-I were detected in the *M.thermoformicicum* strains CSM3, FF1, FF3 and *M.thermoautotrophicum* ΔH. Sequence analysis showed that one FR-I element had been integrated in an almost identical sequence context in the chromosomes of the strains CSM3 and ΔH. Comparison of the FR-I elements from these strains with the one from pFV1 revealed that they consisted of two subfragments, boxI (1118 bp) and boxII (383 bp), with a variable sequential order. Each subfragment was identical on the sequence level with the corresponding one of either FR-I element and was flanked by terminal direct repeats with the consensus sequence A(A/T)ATTT. These results suggest that FR-I represents a mobile element. FR-II was located on both plasmids pFV1 and pFZ1, and on the chromosome of *M.thermoformicicum* strains THF, CSM3 and HN4. Comparison of the nucleotide sequences of the two plasmid-located FR-II copies and the one from the chromosome of strain CSM3 showed that the FR-II segments were approximately 2.5-3.0 kb in size and contained large ORFs that may encode highly related proteins with a yet unknown function.

INTRODUCTION

Genomic rearrangements and mobilization of genetic material are evolutionary processes of prokaryotic and eukaryotic cells that for a part are induced by transposable elements, i.e. insertion sequences (IS) and transposons, due to their ability to move within or between genomes. These mobile elements are also known to be involved in controlling gene expression by providing signals that may enhance or suppress transcription of genes. Being studied in depth within the *Bacteria* (for reviews see: Berg and Howe 1989; Graig and Kleckner 1987), mobile elements have also been described in *Archaea*, a group of phylogenetically related prokaryotes with distinct biochemical and genetic properties (Woese et al. 1990; Zillig 1991)

A variety of diverse IS elements have been characterized in halophilic *Archaea*, one of the three major archaeal groups, and found to be located on plasmids (DasSarma et al. 1983; Pfeifer et al. 1983; Ebert et al. 1984), phages (Schnabel et al. 1984) or chromosomal DNA

(Pfeifer 1986; Hofman et al. 1986). Movement of those transposable elements resulted in a high frequency of DNA-rearrangements causing genome instability in certain halobacteria (Pfeifer et al. 1981; Sapienza et al. 1982). In contrast, among the other two archaeal groups, the sulfur-dependent and the methanogenic organisms, only one IS element, ISM1, was identified in the chromosome of the mesophilic *Methanobrevibacter smithii* (Hamilton and Reeve 1985).

In a previous report we described the isolation of plasmids from several strains of the thermophilic archaeon *Methanobacterium thermoformicicum* (Nölling et al. 1991). The well-characterized low copy number plasmid pFZ1 of strain Z-245 was found to be highly homologous, if not identical to plasmid pFZ2 of strain FTF and partially homologous to plasmid pFV1 from strain THF indicating a close relationship and a common origin of these plasmids. Sequence comparison of pFV1 and pFZ1 revealed a modular genome organization consisting of a highly conserved backbone structure that is interspersed with plasmid-specific elements (Nölling et al. 1992). Two of these elements specify components of restriction-modification (R/M) systems, the GGCC-recognizing *Mth*TI system of pFV1 (Nölling and de Vos 1992a) and the CTAG-speeific *Mth*ZI system carried by pFZ1 (Nölling and de Vos 1992b).

A pFV1-specific element and a region within the conserved backbone structure of both plasmids were found to have homologous counterparts located on the chromosome of several thermophilic, plasmid-free strains of the genus *Methanobacterium*. The analysis of those two homologous regions, described in the present report, showed that at least one shares some structural similarities with IS elements identified in *Bacteria* and other *Archaea*.

METHODS

Bacterial strains and growth conditions

Strains of the genus *Methanobacterium* and *Escherichia coli*, bacteriophages and plasmids used in this study are listed in Table 1. All methanogenic strains were grown as described earlier (Nölling et al. 1991). *E.coli* cultures were cultivated at 37°C in L-broth, if appropriate in the presence of ampicillin (50 μg/ml), and handled as described by Sambrook et al. (1989).

DNA techniques and plasmid constructions

DNA from methanogenic bacteria was isolated with the glass-bead method as described by Nölling et al. (1991). Digestion of chromosomal and plasmid DNA with restriction endonucleases, cloning of DNA fragments, ligation of DNA, transformation of plasmid DNA into *E.coli*, handling of M13 phages and isolation of double and single stranded DNA from M13 phages was carried out essentially as described by Sambrook et al. (1989). DNA of both vector and insert DNA was isolated from agarose gels and purified by the Geneclean procedure (Bio 101, La Jolla, Calif.).

All restriction endonucleases and other DNA-modifying enzymes were obtained from Life Technologies. Oligonucleotides were synthesized with a Biosearch Cyclone DNA synthesizer (New Brunswick Scientific Corp.) at the Netherlands Institute for Dairy Research (NIZO).

All recombinant plasmids were constructed in *E.coli* as described in Table 1. Plasmid

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Property or genotype	Source or reference
M.thermoformicicum		
THF	Harbors plasmid pFV1 (13.5 kb)	DSM3848*; (Zinder and Koch, 1984)
Z-245	Harbors plasmid pFZ1 (11.0 kb)	DSM 3720; (Zhilina and Ilarionov, 1984)
FTF	Harbors plasmid pFZ2 (11.0 kb)	DSM 3012; (Touzel et al., 1988)
FF1, FF3, CSM3, HN4	Plasmid-free	(Nölling et al., 1991, 1993)
SF-4	Plasmid-free	(Yamamoto et al., 1989)
CB12	Plasmid-free	DSM 3664; (Zhao et al., 1986)
M.thermoautotrophicum		
ΔH	Plasmid-free	DSM 1053; (Zeikus and Wolfe, 1972)
Marburg	Harbors plasmid pME2001 (4.4 kb)	DSM 2133; (Fuchs et al., 1978)
E.coli		
DH5α	F *lac*ZΔM15 *rec*A1 *hsd*R17 *sup*E44 Δ(*lac*ZYA-*arg*F)	Life Technologies
TG1	F[*tra*D36 *pro*AB$^+$ *lacI*^{q} *lac*ZΔM15] *sup*E *hsd*Δ5 Δ(*lac*–*pro*AB)	(Gibson, 1984)
Plasmids		
pUC18/pUC19	Ap^r	(Yanisch-Perron et al., 1985)
pUV1	7.4 kb *Sst*I-*Xho*I fragment of pFV1 in pUC19 *Sst*I-*Sal*I, Ap^r	(Nölling and de Vos, 1992)
pUV2	6.1 kb *Sst*I-*Xho*I fragment of pFV1 in pUC19 *Sst*I-*Sal*I, Ap^r	(Nölling and de Vos, 1992)
pUV7	1.5 kb *Acc*I(blunt end)-*Nsi*I fragment of pUV1 in pUC19 *Kpn*I(blunt end)-*Pst*I, Ap^r	This study
pUΔH1	4.0 kb *Eco*RI fragment from *M.thermoautotrophicum* ΔH in pUC19, Ap^r	This study
pUΔH2	1.7 kb *Sca*I-*Pst*I fragment from strain ΔH in pUC19 *Sma*I-*Pst*I, Ap^r	This study
pUCM1	1.4 kb *Eco*RI fragment from *M.thermoformicicum* CSM3 in pUC19, Ap^r	This study
pUCM2	1.2 kb *Eco*RI fragment from strain CSM3 in pUC19, Ap^r	This study
pUCM3	1.6 kb *Sca*I-*Pst*I fragment from strain CSM3 in pUC18 *Sma*I-*Pst*I, Ap^r	This study
pUCM4	1.8 kb *Eco*RI fragment from strain CSM3 in pUC19, Ap^r	This study
pUCM5	1.0 kb fragment amplified from total DNA of strain CSM3 DNA by PCR and cloned into pUC19 *Hind*II, Ap^r	This study
pUCM6	4.5 kb *Eco*RI fragment from strain CSM3 in pUC19, Ap^r	This study

* DSM, German Collection of Microorganisms, Braunschweig, FRG.

pUV7 was constructed by cloning a 2.3-kb *Nsi*I-*Kpn*I fragment of pUV1 into pUC19 digested with *Pst*I and *Kpn*I followed by deletion of a 0.8-kb *Acc*I-*Kpn*I fragment, filling up the ends with Klenow polymerase and religation.

PCR amplification

PCR conditions applied for amplification of a chromosomal fragment were essentially as described earlier (Nölling et al. 1993) using total DNA from *M.thermoformicicum* CSM3 and two oligomers, 5'-CATTTTCACATCTCC and 5'-GTCCTACTGCTTTCAC, complementary to positions 600 to 586 and 1057 to 1072, respectively, of FR-Ic from strain CSM3. The amplified DNA was treated for 30 min at room temperature with Klenow DNA polymerase in the presence of all four dNTPs (2 mM each), followed by an incubation with proteinase K (0.1 mg/ml; Boehringer Mannheim, FRG) and sodium dodecyl sulfate (0.5%) for 20 min at 55°C essentially as reported by Crowe and coworkers (1991). Subsequently, the suspension was subjected to several phenol/chloroform extractions and the DNA was precipitated with ethanol and finally dissolved in TE buffer (Sambrook et al. 1989).

Southern hybridization

DNA was transferred from agarose gels to GeneScreen Plus nylon membranes (NEN-DuPont) by the capillary blot procedure (Sambrook et al. 1989). Hybridization and washing of the filters was carried out according to the membrane manufacturer's protocol. Labelled DNA probes were produced by nick-translation (Sambrook et al. 1989) with [α-^{32}P]dATP (110 TBq/mmol; Amersham). A 1.1-kb *Eco*RI fragment of the pUV7 insert was used to probe for sequences homologous to FR-I (Fig. 1). The FR-II-specific probe consisted of equal amounts of the 2.1-kb *Nar*I-*Sst*I fragment of pUV1 and the 1.2-kb *Sst*I-*Eco*RI fragment of pUV2 (Fig. 1).

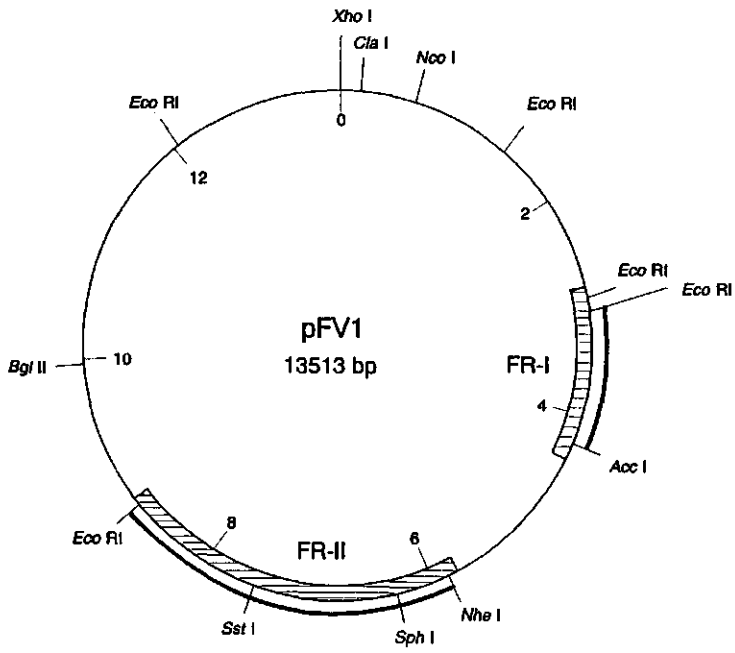


Figure 1. Physical map of plasmid pFV1 from *M.thermoformicicum* THF. Coordinates were taken from Nölling et al. (1993). The regions with homology to the chromosome of other *Methanobacterium* strains, FR-I and FR-II, are indicated by open boxes. The fragments of pFV1 used as FR-I- and FR-II-specific probes are marked by a black bar. Only restriction sites relevant for this study are shown.

Sequence analysis

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al. 1977) with a Sequenase kit (United States Biochemical Corp., Cleveland, Ohio) according to the manufacturer's recommendations, using single stranded DNA prepared from M13mp18/19-based clones (Yanisch-Perron et al. 1985) and vector- and insert-specific primers. Computer analysis of the sequence data was done with the PC/GENE version 5.01 (Genofit, Geneva, Switzerland) while the GCG package version 6.0 (Devereux et al. 1984)

was used to screen the GenBank (release 72.0) and the SwissProt (release 23.0) databases present at the CAOS/CAMM facilities in Nijmegen, The Netherlands.

RESULTS

Sequence homology between plasmid pFV1 and DNA from other methanogens

Total DNA prepared from strains of *M.thermoformicicum* and *M.thermoautotrophicum* were screened for the presence of sequences homologous to plasmid pFV1 from *M.thermoformicicum* THF. In initial hybridization experiments the *Sst*I-*Xho*I inserts of plasmids pUV1 and pUV2 were used as probes (Fig. 1, Table 1). Positive hybridization signals were found under stringent conditions with total DNA isolated from *M.thermoformicicum* strains THF, Z-245, FTF, FF1, FF3, CSM3 and HN4 and with

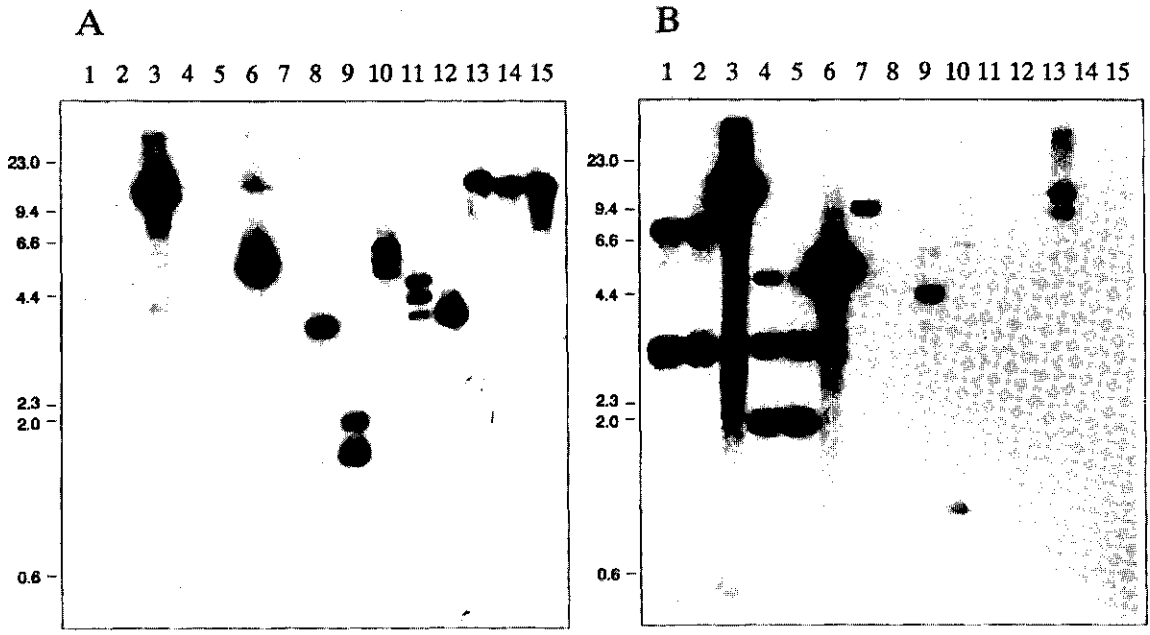


Figure 2. Southern hybridization of total DNA from thermophilic *Methanobacterium* strains with radioactively labelled probes for FR-I (panel A) and FR-II (panel B). Total DNA (about 0.5 μg each) of *M.thermoformicicum* strains Z-245 (lanes 1, 4), FTF (lanes 2, 5), THF (lanes 3, 6), HN4 (lane 7), CSM3 (lanes 9, 10, 13), FF1 (lanes 11, 14) and FF3 (lanes 12, 15), and *M.thermoautotrophicum* ΔH (lane 8) was digested with *Sph*I (lanes 1-3), *Eco*RI (lanes 4-9), *Bam*HI-*Pst*I (lanes 10-12) and *Sal*I-*Sph*I (lanes 13-15), and fractionated by gel electrophoresis on a 0.7% agarose gel. After transfer to a nylon membrane the blots were hybridized under stringent conditions with a 1.1-kb *Eco*RI-*Acc*I fragment of pFV1 (FR-I-specific probe; panel A) and a 3.3-kb *Nar*I-*Eco*RI fragment of pFV1 represented by a 2.1-kb *Nar*I-*Sst*I fragment of pUV1 and a 1.2-kb *Sst*I-*Eco*RI fragment of pUV2 (FR-II-specific probe; panel B). Numbers to the left refer to *Hind*III fragments of λ DNA (in kb) used as size marker.

M.thermoautotrophicum ΔH (Fig. 2), whereas no hybridization was observed with total DNA isolated from *M.thermoformicicum* strains CB12 and SF-4 and *M.thermoautotrophicum* Marburg (not shown). In order to define those homologous regions a series of hybridization experiments was carried out using subfragments of cloned pFV1 DNA as probe. Two distinct regions of plasmid pFV1, designated FR-I (fragment I) and FR-II (fragment II) (see below), were identified that hybridized with DNA from thermophilic *Methanobacterium* strains (Fig. 1; Table 2). The FR-I specific probe showed hybridization with DNA from *M.thermoformicicum* strains CSM3, FF1 and FF3 and *M.thermoautotrophicum* ΔH (Fig. 2A; Table 2), while the probe specific for FR-II hybridized with DNA from *M.thermoformicicum* strains THF, Z-245, FTF, CSM3 and HN4 (Fig. 2B; Table 2).

Table 2. Copy number per chromosome of FR-I and FR-II, respectively, in different strains of the genus *Methanobacterium*

Strain	Copy number	
	FR-I	FR-II
M.thermoformicicum		
THF	4-5 (pFV1)[1]	5-6 (pFV1 and chromosomal DNA)[2]
Z-245	0	4-5 (pFZ1)[1]
FTF	0	4-5 (pFZ2)[1]
CSM3	3	1
FF1	3	0
FF3	3	0
HN4	0	1
CB12	0	0
SF-4	0	0
M.thermoautotrophicum		
ΔH	1[2]	0
Marburg	0	0

[1] estimated on the basis of the plasmid copy number (Nölling et al., 1991).
[2] minimal estimation.

Since three of the hybridizing methanogenic strains were known to contain plasmid DNA, i.e. strains THF, Z-245, and FTF (Nölling et al. 1991), the location of the homologous counterparts of pFV1 was determined. Comparison of the nucleotide sequences of plasmid pFV1 from strain THF and pFZ1 from strain Z-245 (Nölling et al. 1992) had shown that plasmid pFZ1 contained sequences with high similarity to FR-II but not to FR-I (see below). This was confirmed by hybridization with the FR-II probe that showed the expected hybridization pattern of pFZ1 fragments with DNA from strain Z-245 (Fig. 2B, lanes 1 and 4). Since pFZ1 is very similar, if not identical to pFZ2 from strain FTF (Nölling et al. 1991), the latter plasmid should also contain a FR-II-like sequence. Indeed, DNA from both strains Z-245 and FTF showed an identical hybridization pattern (Fig. 2B, lanes 1, 2 and 4, 5).

These results indicate that *M.thermoformicicum* strains Z-245 and FTF do not harbor chromosomal sequences related to plasmid pFV1. Similarly, no chromosomal copies of FR-I seemed to be present in strain THF since only pFV1-derived restriction fragments showed hybridization with the FR-I probe (Fig. 2A, lanes 3, 6). In contrast, hybridization of *Eco*RI-digested total DNA from strain THF with the FR-II-specific probe revealed, besides the expected pFV1-derived fragment (major hybridization species in Fig. 2B, lane 6), additional hybridizing fragments indicating that strain THF harbors chromosomal sequences with similarity to FR-II. Since strains CSM3 and HN4 lack plasmid DNA (unpublished results), the fragments hybridizing to the FR-II probe are part of the chromosomal DNA of these strains. *M.thermoautotrophicum* Marburg, the other known plasmid-containing *Methanobacterium* strain (Meile et al. 1983), appeared to lack sequences with similarity to plasmid pFV1 (Table 2).

Characterization of plasmid and chromosomal copies of FR-I

The exact size and location of FR-I was determined by comparison of the nucleotide sequences of pFV1 and pFZ1. FR-I on pFV1, designated FR-Ia, has a size of 1501 bp (ranging from position 2852 to 4352 according to Nölling et al. 1992), contains terminal, 6 bp direct repeats with the palindromic sequence AAATTT and is flanked by sequences of the conserved backbone structure of both plasmids (Fig. 3 and 4). No sequences with similarity to FR-Ia could be identified in pFZ1. However, plasmid pFZ1 contains a 6-bp sequence (AGATTT) at a location (position 3445 to 3450 according to Nölling et al. 1992) that corresponds with that of FR-Ia in pFV1 (Fig. 4). Since this hexamer resembles the terminal AAATTT repeats of FR-Ia, it may represent a target integration site of FR-I which is duplicated upon insertion. The AT-content of FR-Ia (70.3%) and its flanking regions (58%) was clearly dissimilar suggesting that the origin of FR-Ia differs from that of the rest of the plasmids.

The chromosomal counterparts of FR-I were analyzed in *M.thermoautotrophicum* ΔH and *M.thermoformicicum* CSM3. Two overlapping fragments of *M.thermoautotrophicum* ΔH that hybridized to the FR-I probe, an approximately 4-kb *Eco*RI (Fig. 2A) and a 1.7-kb *Sca*I-*Pst*I fragment (data not shown), were isolated from agarose gel and cloned into *Eco*RI- and *Sma*I-*Pst*I-digested pUC19, respectively, to yield pUΔH1 and pUΔH2 (Table 1; Fig. 3). Both pUΔH1 and pUΔH2 share a common 1-kb *Sca*I-*Eco*RI fragment.

The nucleotide sequence of a 2766-bp *Pst*I fragment included in pUΔH1 and pUΔH2 was determined (EMBL accession no. X69112) (Fig. 5). Comparison of this sequence with that of FR-Ia from plasmid pFV1 revealed that strain ΔH contains a copy of the FR-I sequence, termed FR-Ib, integrated into a region that shares no similarity with plasmid pFV1. The FR-I elements from strains THF and ΔH showed an identical size but different organization: a 383-bp segment, designated boxII, appeared to be located at either side of a 1118-bp segment, designated boxI (Fig. 3 and 5). The distinct organization of the FR-Ia and FR-Ib elements was confirmed by hybridization experiments using the FR-I-specific probe. As the result of the boxI-boxII organization of FR-Ia (Fig. 3), only a single hybridizing *Eco*RI fragment, approximately 5.6 kb in size, was generated (Fig. 2A, lane 6). In contrast, since the FR-Ib element displays a reversed organisation of boxI and boxII (Fig. 3), two hybridizing fragments of approximately 4.0 and 1.3 kb were produced upon cleavage with *Eco*RI (Fig. 2A, lane 8). The sequence of the junction sites of boxI and boxII, AAATTT in pFV1 and ATATTT in strain ΔH, were similar to those found at the ends of FR-I from

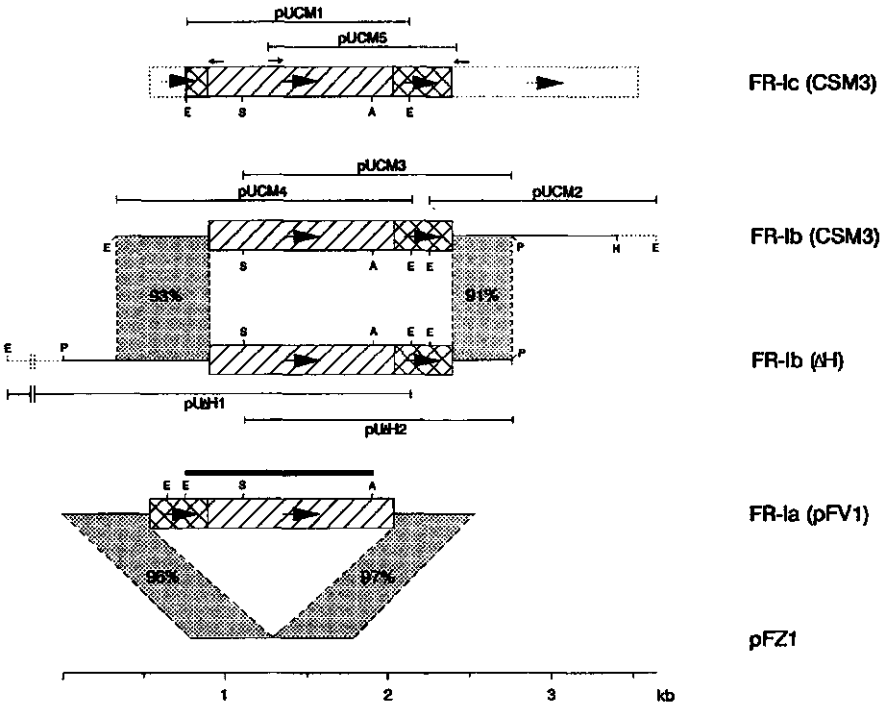


Figure 3. Organization and similarity of different FR-I elements and flanking regions deduced from the nucleotide sequences of the indicated clones described in this study and by Nölling et al. (1993a). Hatched and cross-hatched rectangles in the FR-I elements represent the identical boxI and boxII sequences, respectively. Arrows inside boxI and boxII indicate their orientation. Sequenced regions are denoted by solid lines. FR-Ia represents the FR-I element in plasmid pFV1 from *M.thermoformicicum* THF, whereas FR-Ib and FR-Ic represent the FR-I elements in the chromosome of *M.thermoformicicum* CSM3 and *M.thermoautotrophicum* ΔH, respectively. The presumed extension of FR-Ic is shown by dotted lines. The positions of the FR-I-specific primers used for PCR-amplification of DNA from strain CSM3 are indicated by small arrows. The shaded areas show homologous regions flanking FR-Ib in strains CSM3 and ΔH and those flanking FR-Ia of pFV1 and the corresponding region in pFZ1. The similarity values are indicated in percentages. The restriction fragment of pFV1 used for hybridization as FR-I-specific probe is indicated by a solid bar. Letters denote relevant restriction sites used for the cloning and in hybridization experiments (A, *Acc*I; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sca*I).

either strain, i.e. each boxI, boxII and the complete FR-I element are flanked by similar terminal direct repeats with the consensus sequence A(A/T)ATTT (Fig. 4). Most remarkably, the nucleotide sequences of boxI and boxII from strains THF and ΔH were identical.

M.thermoformicicum CSM3 DNA digested with *Eco*RI and hybridized with the FR-I probe generated two hybridization signals of approximately 1.4 kb and 2.0 kb with different signal intensity (Fig. 2A, lane 9). To clone those fragments, *Eco*RI-digested DNA from strain CSM3 ranging in size from 1-3.5 kb was isolated from agarose gel, ligated to pUC19 and introduced in *E.coli*. Transformants containing DNA that hybridized with the FR-I probe

appeared to harbor one of three plasmids, termed pUCM1, pUCM2 and pUCM3 with inserts of 1.4, 1.3 and 1.8 kb, respectively, (Fig. 3; Table 1). The similar sized inserts cloned in pUCM1 and pUCM2 may explain the differences in signal intensity obtained with the FR-I probe (Fig. 2A). A fourth clone, pUCM4, was constructed by cloning a 1.6-kb *Sca*I-*Pst*I fragment from strain CSM3 which overlaps with the inserts of pUCM2 and pUCM3, into pUC18 digested with *Sma*I-*Pst*I (Fig. 3).

Nucleotide sequence analysis of the pUCM3- and pUCM4 inserts and of a 1.2-kb *Eco*RI-*Hind*III fragment of the pUCM2 insert yielded a continuous stretch of 3055 bp. Comparison of this sequence to that of FR-I from *M.thermoformicicum* THF and *M.thermoautotrophicum* ΔH revealed that *M.thermoformicicum* CSM3 also contains a complete FR-Ib-type element (EMBL accession no. X69113) analogous to the one found in strain ΔH (Fig. 3). Direct repeats with the sequence A(A/T)ATTT were flanking either end of boxI, boxII and the entire FR-I element (Fig. 4). Again, the CSM3-derived boxI and boxII were identical on the sequence level with those from strains THF and ΔH.

pFZ1	G C A C C C C A T G A G A T T T T A A T A T G A A A	
pFV1	C C A C C C T A G G A A A T T T C T C T T T A C A G	left end FR-Ia (boxII)
pFV1	G A A A T C C G T T A A A T T T T A A T A T G A A A	right end FR-Ia (boxI)
ΔH	T G A T G T T G A G A A A T T T G A A A T T A A A A	left end FR-Ib (boxI)
ΔH	A A C A A T A G T T A T A T T T G A A T T T G T G G	right end FR-Ib (boxII)
CSM3	C G A T G T T G A G A A A T T T G A A A T T A A A A	left end FR-Ib (boxI)
CSM3	A A C A A T A G T T A T A T T T G A A T T T G T G G	right end FR-Ib (boxII)

pFV1	A A C A A T A G T T A T A T T T G A A A T T A A A A	junction site boxII/boxI in FR-Ia
ΔH	G A A A T C C G T T A A A T T T C T C T T T A C A G	junction site boxI/boxII in FR-Ib
CSM3	G A A A T C C G T T A A A T T T C T C T T T A C A G	junction site boxI/boxII in FR-Ib
CSM3	A A C A A T A G T T A T A T T T G A A A T T A A A A	junction site boxII/boxI in FR-Ic
CSM3	G A A A T C C G T T A A A T T T C T C T T T A C A G	junction site boxI/boxII in FR-Ic

Figure 4. Conserved terminal hexanucleotides in the FR-I elements and boxI-boxII junction sites, respectively. The upper part shows an alignment of the terminal sequences of FR-I elements derived from plasmid pFV1 *M.thermoformicicum* THF (FR-Ia) and the corresponding region of plasmid pFZ1 from *M.thermoformicicum* Z-245, and from the chromosome of *M.thermoformicicum* CSM3 (FR-Ib and FR-Ic) and *M.thermoautotrophicum* ΔH (FR-Ib). The boxI-boxII junction regions of either FR-I element is shown in the lower alignment. The consensus hexanucleotide A(A/T)ATTT is boxed.

Similar to the situation observed in plasmid pFV1 the flanking regions of FR-I in strain CSM3 and ΔH also differed in AT-content from that of FR-I, i.e. 45.9% versus 70.3%, respectively. This suggests that these elements have been acquired from another source and integrated into the chromosome. Alignment of an approximately 0.5 kb sequence stretch flanking FR-I from the strains CSM3 and ΔH revealed extensive similarity (>90% identity)

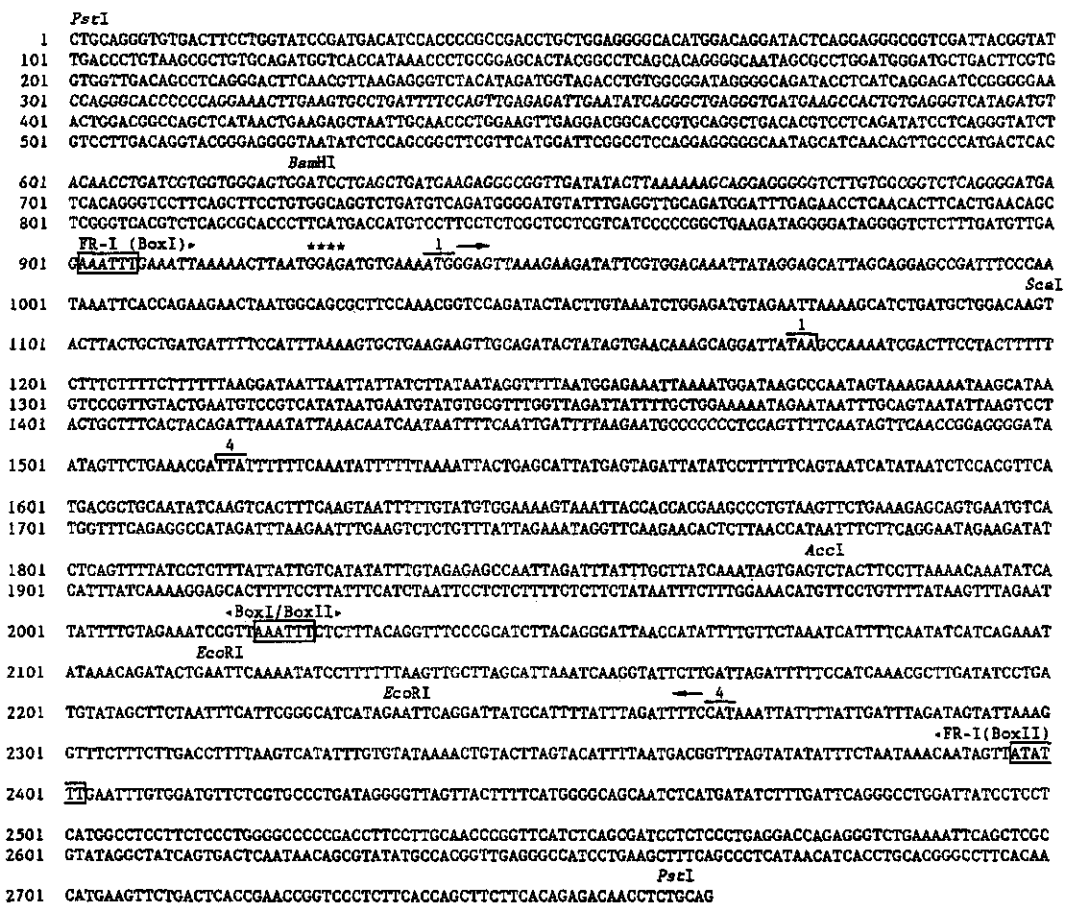


Figure 5. Nucleotide sequence of the 2766-bp *Pst*I fragment from *M.thermoautotrophicum* ΔH including FR-Ib and its flanking regions. Boxed nucleotides represent the consensus hexanucleotide located at the ends of FR-I and the junction site of boxI and boxII. The initiation and termination codons of two ORFs (1 and 4) of FR-Ib are marked. Asterisks denote a potential ribosome binding site.

(Fig. 3). Since strains ΔH and CSM3 are closely related organisms (Nölling et al. 1993b) this finding suggests that the FR-Ib element is located at identical positions within the same sequence context in the chromosomes of strain ΔH and CSM3.

A different result was obtained from sequence analysis of the 1.4-kb *Eco*RI fragment cloned in pUCM1. It contained the complete boxI of FR-I which seemed to be flanked at both sites by one copy of boxII (Fig. 3). The junction sites of boxI and boxII showed the consensus sequence A(A/T)ATTT (Fig. 4). Since the cloned *Eco*RI fragment contained only about 100 bp of the adjacent copies of boxII no analysis of the flanking chromosomal sequences was possible. Nevertheless, the sequence of the pUCM1 insert suggests that, in addition to a FR-Ib, strain CSM3 contains another type of FR-I, designated FR-Ic, consisting

of either two copies of boxII and one copy of boxI or representing a tandem repeat of either FR-Ia or FR-Ib (Fig. 3). To discriminate between these possibilities, we analyzed a PCR-amplified DNA fragment that was produced from total DNA of strain CSM3 using two primers specific for boxI of FR-I (see Fig. 3). Sequence analysis of the approximately 1-kb PCR product, which was cloned into *Hind*II-linearized pUC19 to yield pUCM5 (Fig. 3; Table 1), revealed that it comprise a complete boxII flanked by sequences homologous to boxI. This result suggests that FR-Ic may consist of at least two continuous copies of either FR-Ia or FR-Ib (Fig. 3).

Coding capacity of the FR-I elements

The FR-I elements were inspected for the presence of open reading frames (ORFs) with more than 200 bp and appropriate start codons (ATG, GTG or TTG). The boxI of either FR-I element contained an identical ORF, designated ORF1, with a size of 237 bp ranging from position 3264 to 3500 in FR-Ia of strain THF (Nölling et al. 1992) and from position 937 to 1173 in FR-Ib of strain ΔH (Fig. 5 and 6). Two additional ORFs were found in FR-Ia of pFV1. One of these, ORF2, was 294 bp in size (ranging from position 3100 to 2809) and part of boxII, while the other one, ORF3, was 258 bp in length (from position 4103 to 3848) and located in boxI (Fig. 6). In FR-Ib of the strains ΔH and CSM3, the different organization of boxI and boxII resulted in a fusion of ORF2 and ORF3 yielding ORF4 with a size of 750 bp (ranging from position 2268 to 1519 in strain ΔH; Fig. 5 and 6). However, except for ORF1, none of the ORFs were preceded by a sequence with similarity to a consensus methanogenic ribosome binding site (Brown et al. 1989) (Fig. 5). A TFASTA search (Pearson et al., 1988) of the nucleotide sequence and the deduced amino acid

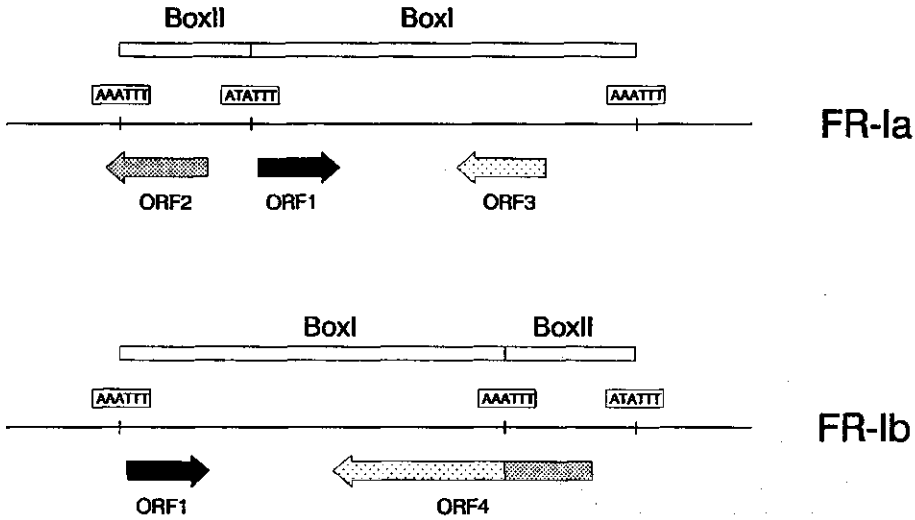


Figure 6. Coding capacity of FR-Ia from plasmid pFV1 of *M.thermoformicicum* THF and FR-Ib from either *M.thermoautotrophicum* ΔH or *M.thermoformicicum* CSM3. Identical coding regions are indicated by a similar shading.

sequences of possible gene products of the FR-I elements with those stored in the databases revealed no significant similarity.

Copy number of FR-I elements

Since the FR-I elements were exclusively plasmid-located in *M.thermoformicicum* THF, their copy number correspond with that of pFV1 which is present in approximately 4 to 5 copies per chromosome in this strain (Nölling et al. 1991) (Table 2). To determine the copy number of FR-I elements contained in the chromosome of *Methanobacterium* strains, total DNA of these strains was digested with restriction endonucleases that have no cleavage sites in FR-I, and subsequently hybridized with the FR-I-specific probe. Hybridization of *Sal*I-*Sph*I cleaved DNA prepared from *M.thermoformicicum* strains CSM3, FF1 and FF3 yielded a single hybridization signal with comparable intensity (Fig. 2A, lanes 13-15). When DNA of the same strains was digested with *Bam*HI and *Pst*I, we observed three hybridization signals with the DNA from strains CSM3 and FF1 and only one with that from strain FF3 (Fig. 2A, lanes 10-12). The obtained hybridization patterns and the signal intensities indicate that strains CSM3, FF1 and FF3 each harbor at least three copies of a FR-I element which are located on *Bam*HI-*Pst*I fragments that differ in size in strains CSM3 and FF1 but are similar sized in strain FF3 (Table 2). Moreover, these multiple FR-I elements seem to be clustered in the strains CSM3, FF1 and FF3 on an approximately 15 kb sized *Sal*I-*Sph*I fragment (Fig. 2A, lanes 13-15).

In the case of *M.thermoautotrophicum* ΔH, a single hybridization signal was observed with *Bam*HI-digested DNA (data not shown) and two signals with DNA cleaved by *Eco*RI (FR-I contains two *Eco*RI sites; Fig. 2A, lane 8). These results would indicate that strain ΔH contains a single copy of FR-I. However, the comparable signal intensity of the major hybridization species of *Eco*RI-digested ΔH DNA (Fig. 2A, lane 8) and that derived from *Bam*HI-*Pst*I cleaved FF3 DNA which represents three copies (Fig. 2A, lane 12), suggests that strain ΔH may harbor more than one FR-I element (Table 2).

Characterization of FR-II

Comparison of the nucleotide sequences of plasmid pFV1 and pFZ1 (Nölling et al. 1992) had shown that plasmid pFZ1 contains a region homologous to the 3.3-kb *Nar*I-*Eco*RI fragment of pFV1 which includes FR-II. Therefore, plasmid pFZ1 also carries a copy of FR-II (designated FR-II'). Due to the high interplasmid similarity of the FR-II containing regions, however, determination of the exact size of FR-II and FR-II' was not possible. To allow for a better comparison, the chromosomal FR-II element of *M.thermoformicicum* CSM3 was analyzed. For this purpose, we first cloned a 4.5-kb *Eco*RI fragment of strain CSM3 that hybridized with the FR-II probe (Fig. 2B, lane 9) to yield pUCM6 (Table 1). Sequence analysis of a part of the pUCM6 insert generated a continuous stretch of 2676-bp (EMBL accession no. X69114). Alignment of this sequence with that of the FR-II-containing plasmid regions revealed a 2.5-kb fragment with high similarity. This chromosomal segment of strain CSM3 was designated cFR-II (Fig. 7).

The cFR-II segment comprised 2542 bp (from position 47 to 2568 of the analyzed pUCM6 sequence) and had a similar size as FR-II' from plasmid pFZ1 (2510 bp; position 4750 to 7258 according to Nölling et al. 1992). FR-II from plasmid pFV1 contained, however, a direct repeat of two identical 524-bp sequences (DR) that enlarged its size to 3034 bp

(position 5687 to 8719 according to Nölling et al., 1992).

Both FR-II' and cFR-II contained ORFs of a comparable size, ORF5' (position 5295 to 6791) and cORF5 (position 730 to 2142) (Fig. 7), respectively, that could code for proteins with a calculated molecular mass of respectively 54.1 and 51.6 kDa. Due to the sequence duplication in pFV1, however, two partially identical ORFs, ORF5a (position 6214 to 6221) and ORF5b (position 6890 to 6895), were generated in FR-II that could code for proteins with a calculated molecular mass of 24.7 and 49.6 kDa, respectively (Fig. 7). As a consequence, the C-terminal part of the ORF5a protein and the N-terminal part of the ORF5b protein would share a block of 170 identical amino acids. The plasmid-located ORF5a of pFV1 and ORF5 of pFZ1 are followed by two additional, smaller ORFs, ORF6 (position 8354 to 8719) and ORF6' (position 6831 to 7259), respectively (Fig. 7). The capacity to code for large and highly similar proteins is the most conspicuous feature of the FR-II segments. Alignment of the amino acid sequences deduced from ORF5a/b, ORF5' and cORF5 revealed 72-84% identical residues. Interestingly, cORF5 of strain CSM3 is flanked by sequence blocks which share no homology with the plasmid-located FR-II segments (Fig. 7). This particular conservation of cORF5 and its plasmid-derived counterparts may indicate

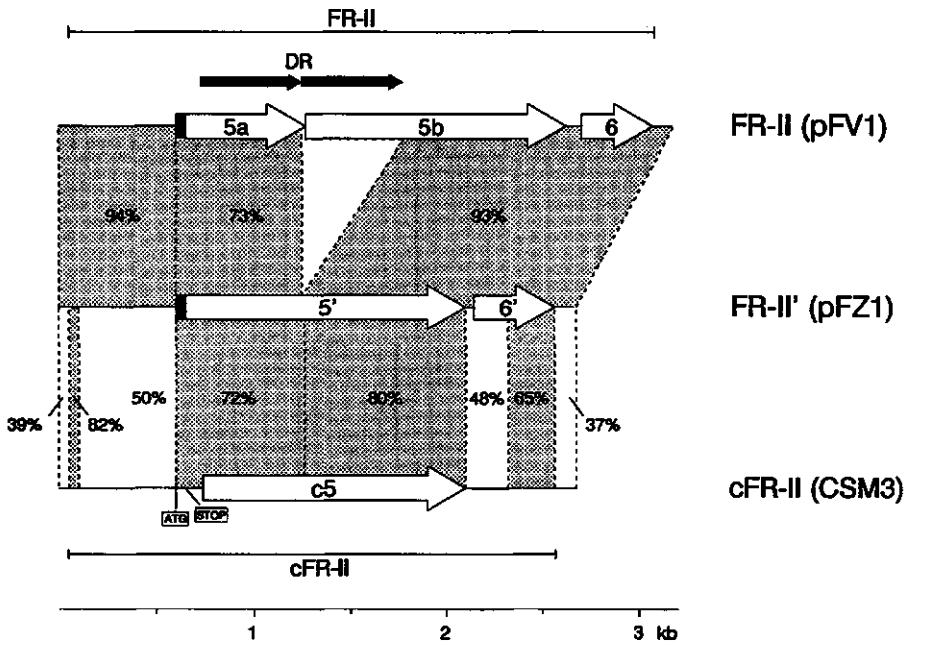


Figure 7. Organization and similarity of the FR-II segments derived from plasmid pFV1 and pFZ1 of *M.thermoformicicum* strains THF and Z-245, respectively, and the chromosome of strain CSM3. ORFs are represented by open arrows whereas the direct repeat (DR) found in pFV1 is indicated by black arrows. The 5'-region of ORF5a and ORF5' that could code for a putative signal peptide is indicated by a black block. The position of a stop codon that would result in a N-terminal truncated cORF5 protein of strain CSM3 is shown. Regions with high degree of similarity are shaded and homology values are given in percentages.

their expression into proteins with yet unknown function. A database search failed, however, to detect significant similarities with the nucleotide sequence or the deduced amino acid sequence of either of the FR-II segments.

The ORF5a protein of pFV1 and the ORF5 protein of pFZ1 contained a stretch of 24 amino acids with structural similarities to a prokaryotic signal sequence (Gierasch 1989): a N-terminal located block of four polar and one positively charged residue followed by a hydrophobic core of 13 residues and a putative signal peptidase cleavage site (for ORF5a: NH_2-MRNGT*LVLAALLTLFVLA*GSSSA↓*A*D*V*G*I*E*L*DKN...; charged and hydrophobic residues are underlined and in italics, respectively; the arrow indicates the putative signal peptidase cleavage site). The presence of this putative leader peptide suggests that the proteins may be translocated through or associated with the cell membrane. The deduced product of cORF5 of strain CSM3, however, lacks this sequence since a stop codon at position 688 would result in a protein which is, compared to the corresponding plasmid-encoded proteins, N-terminally truncated by 33 amino acids (Fig. 7).

Inspection for appropriate expression signals suggested that the ORF5a/ORF5b region of plasmid pFV1 may form a transcriptional unit since a sequence stretch with considerable similarity to a typical archaeal promoter signal (Hain et al. 1991) could be identified upstream ORF5a, whereas ORF5b is immediately followed by a T-rich stretch that resembled a conserved archaeal transcription termination signal (Brown et al. 1989). Both ORFs are preceded by a potential ribosome binding site (Brown et al. 1989). The putative transcription and translation signals identified for ORF5' of pFZ1 were almost identical to those found for ORF5a/ORF5b, i.e. ORF5' displays the promoter sequence and the ribosome binding site of ORF5a and the terminator signal of ORF5b. Similar transcription signals could be identified for the chromosomal cORF5 although it is devoid of a appropriate ribosome binding site.

DISCUSSION

The high similarity of the archaeal plasmids pFV1, pFZ1 and pFZ2 from *M.thermoformicicum* strains THF, Z-245 and FTF, respectively, suggested by hybridization studies (Nölling et al. 1991) and partially confirmed by nucleotide sequence analysis (Nölling et al., 1992), implies that these extrachromosomal elements belong to a family of related plasmids and probably have been derived from a common ancestor. Whereas the major part of the plasmids consist of unique sequences which showed no homology to chromosomal DNA of the methanogenic strains used in this study, two regions of plasmid pFV1 could be identified that have chromosome-located counterparts in several thermophilic strains of the genus *Methanobacterium*. One of those plasmid-derived regions, FR-I, is characteristic for plasmid pFV1, whereas the other one, FR-II, is also part of plasmids pFZ1 and pFZ2.

FR-I consists of two sequence blocks, boxI and boxII, which occur in a variable sequential order. Based on their organization three different types of FR-I could be identified (Fig. 3): (i) FR-Ia, which is part of plasmid pFV1, displays a boxII-boxI organization; (ii) FR-Ib, isolated from the chromosome of *M.thermoautotrophicum* ΔH and *M.thermoformicicum* CSM3, with the order boxI-boxII; and (iii) FR-Ic from strain CSM3 consisting of tandemly repeated copies of either FR-Ia or FR-Ib. In spite of their different organization, boxI and boxII of each element have the same orientation (Fig. 3) and an identical nucleotide sequence. FR-Ia and FR-Ib are 1501 bp in size and exhibit defined short terminal direct repeats with the consensus sequence A(A/T)ATTT. Moreover, this consensus sequence

defines the boundaries of boxI and boxII (Fig. 4). The observed different organization of boxI and boxII could well be the result of recombination. It may be suggested that either the A(A/T)ATTT sequence is involved in the rearrangement of boxI and boxII or, alternatively, that the elements containing only a single boxI, i.e. FR-Ia and FR-Ib, represent deletion derivatives of a FR-Ic-like element. The single FR-Ib element found in strain CSM3, for instance, could have been generated from a FR-Ic-type element consisting of two repeated FR-Ib copies by homologous recombination between those repeated elements.

An exceptional property of the FR-I elements is their complete conservation on the sequence level as found for four independently isolated copies from three different *Methanobacterium* strains. A similar observation was reported for the halobacterial insertion sequences ISH1 (Simsek et al. 1982; Pfeifer and Ghahraman 1991) and ISH2 (DasSarma et al. 1983; Pfeifer and Ghahraman 1991). Analogous to the 520-bp ISH2 element, also the FR-I elements have low coding capacity, indicating that other characteristics than the encoded gene products might be responsible for the conservation on the nucleotide level. One may argue that the examined strains have recently acquired these elements, and therefore base variations are unlikely to have occurred during this timespan. However, all methanogenic strains that contain at least one copy of an FR-I type were isolated from geographically very different locations during a period of two decades.

It is tempting to speculate that FR-I represents a new archaeal insertion element. Several observations would support this hypothesis. (i) The FR-I elements exist in multiple genomic copies. At least two of the elements, FR-Ia and FR-Ib, are part of a different sequence context as revealed from sequence comparison of the FR-I flanking regions. Since the flanking regions showed no sequence similarity to the FR-I elements, integration of the elements is unlikely to have occurred by homologous recombination. (ii) Each FR-I element contains terminal direct repeats with the consensus sequence A(A/T)ATTT. This direct repeat may either be part of the element or may represent a target site duplication generated upon integration of FR-I. The latter possibility would be supported by the fact that plasmid pFZ1 from strain Z-245 contains a single copy of a similar hexanucleotide, AGATTT, at the position where FR-I is integrated in plasmid pFV1 (Fig. 5). (iii) The FR-I elements contain ORFs that may encode proteins involved in movement of the elements. Since, however, rearrangements of boxI and boxII affect the coding capacity, the major 750-bp ORF is only found in FR-Ib and FR-Ic (Fig. 6).

In contrast to most of the IS elements known in *Archaea* and *Bacteria*, FR-I lacks terminal inverted repeats. However, transposition does not depend on the presence of terminal inverted repeats as was found for several transposons from gram-positive bacteria, such as Tn*554* (Murphy and Lõfdahl 1984) or Tn*5276* (Rauch and de Vos 1992). Similarly, no terminal inverted repeats were found for the (putative) insertion element ISH1.8 associated with phage ΦH of the archaeon *Halobacterium salinarium* (Schnabel et al. 1984). The latter authors concluded that the insertion mechanism of those elements is different from that of IS elements with terminal inverted repeats. This could also apply to FR-I. The conclusion, however, that FR-I is a novel type of IS element will only be justified by the observation of its transposition.

A second region of plasmid pFV1, FR-II, was found to have homologous counterparts in the chromosome of *M.thermoformicicum* strains CSM3 and HN4. Furthermore, a chromosomal copy of FR-II seems also to be present in strain THF, the host of plasmid pFV1. In contrast to FR-I, FR-II is an integral part of plasmid pFV1 and pFZ1 and possibly also of the postulated common ancestor of those related plasmids. The presence of FR-II in a restricted number of strains on either plasmid or chromosomal DNA suggests that it belongs

to a similar class of elements as FR-I. However, compared to FR-I, FR-II differs in several aspects: (i) FR-II is also part of the plasmids pFZ1 and pFZ2 of the strains Z-245 and FTF, respectively; (ii) it is less frequently distributed among other thermophilic *Methanobacterium* strains (Table 2); (iii) only a single copy of FR-II seems to be located on chromosomal DNA (Table 2); (iv) FR-II is less conserved on the nucleotide level; (v) there is no evidence for target site duplication; (vi) it displays a higher coding capacity. The most remarkable characteristic shared by the analyzed two plasmid-located FR-II segments and the chromosomal one from strain CSM3 is the presence of a large coding region that may direct the synthesis of highly similar proteins. The high conservation of the FR-II coding regions may indicate that they encode proteins with yet unknown function.

The distribution of FR-I among thermophilic *Methanobacterium* strains strongly correlates with their phylogenetic position. As was shown by DNA-DNA hybridization studies and 16S rRNA analysis, all FR-I carrying strains, i.e. *M.thermoformicicum* strains THF, CSM3, FF1 and FF3, and *M.thermoautotrophicum* ΔH, belong to the same species (Touzel et al. 1991; Nölling et al. 1993). We can imagine two explanations for this finding. Either FR-I was already part of the common ancestor of those strains or its dissemination is the result of horizontal gene transfer. The fact that the strains have been isolated from geographically very different locations, would argue for the first possibility. If so, this would imply that the strains Z-245 and FTF, which belong to the same species as the FR-I harboring strains (Touzel et al. 1992; Nölling et al. 1993) but lack FR-I, have lost this element. On the other hand, lateral gene transfer of FR-I may have occurred analogous to dissemination of insertion elements in *Bacteria* via plasmids or phages. In this case plasmid pFV1 from strain THF may have served as vector in mobilizing FR-I from one strain to another.

REFERENCES

Berg EG, Howe DE (eds.) (1991) Mobile DNA, American Society for Microbiology, Washington, DC

Brown JW, Daniels CJ, Reeve JN (1989) Gene structure, organization, and expression in archaebacteria. Crit Rev Microbiol 16:287-337

Craig NL, Kleckner N (1987) Transposition and site-specific recombination. In: *Escherichia coli* and *Salmonella typhimurium*, p. 1054-1070 (eds. FC Neidhardt), American Society for Microbiology, Washington DC

Crowe JS, Cooper HJ, Smith MA, Sims MJ, Parker D, Gewert D (1991) Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acids Res 19:184

DasSarma S, RajBhandary UL, Khorana HG (1983) High frequency spontaneous mutations in the bacterio-opsin gene in *Halobacterium halobium* is mediated by transposable elements. Proc Natl Acad Sci USA 80:2201-2205

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Res 12:387-395

Ebert K, Goebel W, Pfeifer F (1984) Homologies between extrachromosomal DNA populations of *Halobacterium halobium* and four new halobacterial isolates. Mol Gen Genet 194:91-97

Fuchs G, Stupperich E, Thauer RK (1978) Acetate assimilation and synthesis of alanine, aspartate and glutamate in *Methanobacterium thermoautotrophicum*. Arch Microbiol 117:61-66

Gierasch LM (1989) Signal sequences. Biochemistry 28:923-930

Hain J, Reiter W-D, Hüdepohl U, Zillig W (1992) Elements of an archaeal promoter defined by mutational analysis. Nucleic Acids Res. 20:5423-5428

Hamilton PT, Reeve JN (1985) Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. Mol Gen Genet 200:47-59

Hofman JD, Schalkwyk LC, Doolittle WF (1986) ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaebacterium, *Halobacterium volcanii*. Nucleic Acids Res 14:6983-6998

Meile L, Kiener A, Leisinger T (1983) A plasmid in the archaebacterium *Methanobacterium thermoautotrophicum*. Mol Gen Genet 191:480-484

Murphy E, Löfdahl S (1984) Transposition of Tn*554* does not generate a target duplication. Nature (London) 307:292-294

Nölling J, Frijlink M, de Vos WM (1991) Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. J Gen Microbiol 137:1981-1986 (Chapter 4)

Nölling J, de Vos WM (1992a) Characterization of the archaeal, plasmid-encoded type II restriction-modification system *Mth*TI of *Methanobacterium thermoformicicum* strain THF: homology to the bacterial *Ngo*PII-system from *Neisseria gonorrhoeae*. J Bacteriol 174:5719-5726 (Chapter 7)

Nölling J, de Vos WM (1992b) Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene. Nucleic Acids Res 20:5047-5052 (Chapter 8)

Nölling J, van Eeden FJM, Eggen RIL, de Vos WM (1992) Modular organization of related archaeal plasmids encoding different restriction-modification systems in *Methanobacterium thermoformicicum*. Nucleic Acids Res 20, 6501-6507 (Chapter 5)

Nölling J, Hahn D, Ludwig W, de Vos WM (1993) Phylogenetic analysis of several thermophilic *Methanobacterium* sp.: evidence for a formate utilizing ancestor. System Appl Microbiol, in press (Chapter 2)

Pearson DJ, Lipman WR (1988) Improved tools for the biological sequence analysis. Proc Natl Acad Sci USA 85:2444-2448

Pfeifer F, Weidinger G, Goebel W (1981) Genetic variability in *Halobacterium halobium*. J Bacteriol 145:375-381

Pfeifer F, Betlach M, Martienssen R, Friedman J, Boyer H (1983) Transposable elements of *Halobacterium halobium*. Mol Gen Genet 191:182-188

Pfeifer F (1986) Insertion elements and genome organization of *Halobacterium halobium*. System Appl Microbiol 7:36-40

Pfeifer F, Ghahraman P (1991) The halobacterial insertion element ISH28. Nucleic Acids Res 19:5788

Rauch PJG, de Vos WM (1992) Characterization of the novel nisin-sucrose conjugative transposon Tn*5276* and its insertion in *Lactococcus lactis*. J Bacteriol 174:1280-1287

Sapienza C, Rose MR, Doolittle WF (1982) High-frequency genomic rearrangements involving archaebacterial repeat sequence elements. Nature (London) 299:182-185

Schnabel H, Palm P, Dick K, Grampp B (1984) Sequence analysis of the insertion element ISH1.8 and of associated structural changes in the genome of phage ΦH of the archaebacterium *Halobacterium halobium*. EMBO J 3:1717-1722

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain termination inhibitors. Proc Natl Acad Sci USA 74:5463-5467

Simsek M., DasSarma S., RajBhandary UL, Khorana HG (1982) A transposable element from *Halobacterium halobium* which inactivates the bacteriorhodopsin gene. Proc Natl Acad Sci USA 79:7268-7273

Touzel JP, Petroff D, Maestrojuan GM, Prensier G, Albagnac C (1988) Isolation and characterization of a thermophilic *Methanobacterium* able to use formate, the strain FTF. Arch Microbiol 149:291-296

Touzel JP, Conway de Macario E, Nölling J, de Vos WM, Zhilina T, Lysneko AM (1992) DNA

relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int J Syst Bacteriol 42:408-411

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene 33:103-119

Yamamoto K, Tachibana A, Dhavises G, Tanaka T, Taniguchi M, Oi S (1989) Characterization of a thermophilic formate-utilizing methanogen, *Methanobacterium thermoformicicum* strain SF-4. Agr Biol Chem 53:533-543

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576-4579

Zeikus JG, Wolfe RS (1972) *Methanobacterium thermoautotrophicum* sp. n., an anaerobic, autotrophic, extreme thermophile. J Bacteriol 109:707-713

Zhao Y, Zhang H, Boone D, Mah RA (1986) Isolation and characterization of a fast-growing, thermophilic *Methanobacterium* species. Appl Environ Microbiol. 52:1227-1229

Zhilina TN, Ilarionov SV (1984). Characteristics of formate-assimilating methane bacteria and description of *Methanobacterium thermoformicicum* sp. nov. Microbiol 53:647-651 (English translation of Mikrobiologyia 53:785-790)

Zillig W (1991) Comparative biochemistry of *Archaea* and *Bacteria*. Current Opinion in Genetics and Development 1:544-551

Zinder SH, Koch M (1984) Non-acetoclastic methanogenesis from acetate: acetate oxidation by a thermophilic syntrophic coculture. Arch Microbiol 138:263-272

Characterization of the archaeal, plasmid-encoded type II restriction-modification system *Mth*TI from *Methanobacterium thermoformicicum* THF: homology to the bacterial *Ngo*PII system from *Neisseria gonorrhoeae*

Jörk Nölling and Willem M. de Vos

Journal of Bacteriology **174**, 5719-5726.

JOURNAL OF BACTERIOLOGY, Sept. 1992, p. 5719–5726
0021-9193/92/175719-08$02.00/0

Vol. 174, No. 17

Characterization of the Archaeal, Plasmid-Encoded Type II Restriction-Modification System *Mth*TI from *Methanobacterium thermoformicicum* THF: Homology to the Bacterial *Ngo*PII System from *Neisseria gonorrhoeae*

JÖRK NÖLLING AND WILLEM M. DE VOS*

Bacterial Genetics Group, Department of Microbiology, Agricultural University of Wageningen, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

Received 16 April 1992/Accepted 1 July 1992

A restriction-modification system, designated *Mth*TI, was localized on plasmid pFV1 from the thermophilic archaeon *Methanobacterium thermoformicicum* THF. The *Mth*TI system is a new member of the family of GGCC-recognizing restriction-modification systems. Functional expression of the archaeal *Mth*TI genes was obtained in *Escherichia coli*. The *mthTIR* and *mthTIM* genes are 843 and 990 bp in size and code for proteins of 281 (32,102 Da) and 330 (37,360 Da) amino acids, respectively. The deduced amino acid sequence of M.*Mth*TI showed high similarity with that of the isospecific methyltransferases M.*Ngo*PII and M.*Hae*III. In addition, extensive sequence similarity on the amino acid level was observed for the endonucleases R.*Mth*TI and R.*Ngo*PII. Moreover, the endonuclease and methyltransferase genes of the thermophilic *Mth*TI system and those of the *Neisseria gonorrhoeae Ngo*PII system show identical organizations and high (54.5%) nucleotide identity. This finding suggests horizontal transfer of restriction-modification systems between members of the domains *Bacteria* and *Archaea*.

Type II restriction-modification (R/M) systems are composed of two enzymatic activities, an endodeoxyribonuclease (ENase) and a DNA methyltransferase (MTase), which recognize the same specific DNA sequence (44). It has been proposed that a major function of R/M systems is the protection of resident DNA against contamination by sequences of foreign origin (16). A great variety of R/M systems have been identified in prokaryotes (13, 30) and appeared to be valuable tools for studying evolution at the gene and protein levels (44).

Genetic characterization of R/M systems showed that the ENase and MTase genes are usually located in close proximity to each other, although their relative orientations may differ (44). Comparison of ENases and MTases from sequenced R/M systems revealed that ENases share no protein similarity with isospecific MTases. In addition, ENases generally show no homology with other ENases (18), with the exception of the pairs R.*Eco*RI-R.*Rsr*I (36) and R.*Bsu*FI-R.*Msp*I (12). In contrast, considerable similarities between various MTases were found at the amino acid level, especially between those generating 5-methylcytosine (m^5C-MTases) (15, 19, 26, 37). Among R/M systems with m^5C-MTases, those that recognize the target sequence GGCC are widely distributed (30). Representatives that have been sequenced include *Hae*III (35), *Ngo*PII (37, 38), *Bsu*RI (14), and *Bsp*RI (27). In addition, the sequences from three GGCC-specific MTases from *Bacillus subtilis* phages SPR, ϕ3T, and $\rho 11_s$ have been determined (1, 41). This molecular characterization of GGCC-specific R/M systems allowed the identification of conserved sequence motifs such as the target-recognizing domains in the MTases and the study of the evolution of these proteins (1, 17–19, 43).

All known sequenced R/M systems originate from members of the domain *Bacteria*; no sequences of ENases or MTases have been reported from members of the domain *Archaea*, a fundamentally different group of prokaryotes with a unique phylogenetic position (46). Nevertheless, R/M systems have been detected in representatives of all archaeal groups, including halobacteria (25), thermoacidophiles (28), and several methanogens (20, 33, 39).

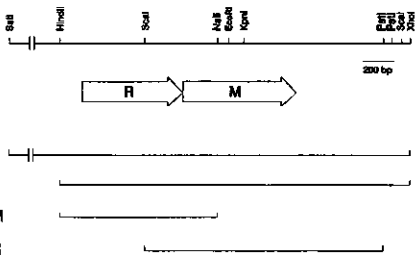
In this communication, we describe the characterization of a GGCC-recognizing R/M system, designated *Mth*TI, from the thermophilic archaeon *Methanobacterium thermoformicicum* THF. This R/M system was initially detected during molecular characterization of a new plasmid, pFV1, that was recently identified in this strain (23). Unexpectedly, the gene and deduced amino acid sequences of the *Mth*TI system showed significant sequence identity with those of the *Ngo*PII system of *Neisseria gonorrhoeae*, suggesting horizontal gene transfer between members of the *Archaea* and *Bacteria*.

MATERIALS AND METHODS

Bacterial strains and plasmids. *M. thermoformicicum* THF (DSM 3848) was cultivated as described previously (23). *Escherichia coli* JM83 (47) and TG1 (9) were used for subcloning and propagation of plasmids pUC18 and pUC19 (47) and phages M13mp18 and M13mp19 (47), respectively. Expression of the *mthTIR* and *mthTIM* genes was analyzed in *E. coli* HB101 (2). *E. coli* strains were grown at 37°C in L broth, if appropriate in the presence of ampicillin (50 μg/ml), and handled according to established procedures (31).

The entire 13.5-kb plasmid pFV1 from *M. thermoformicicum* THF (24) was cloned into a λ GEM-12 vector (Promega Biotec, Madison, Wis.) by using the single *Xho*I site of pFV1, yielding phage λ GFV1 (24). For construction of plasmids pUV1 and pUV2, two *Xho*I-*Sst*I fragments, 7.4 and 6.1 kb in size, of λ GFV1 were subcloned into pUC19 digested with *Sal*I and *Sst*I. Plasmid pUV1 was used as a substrate in endonuclease and methylase assays. Plasmid

* Corresponding author.



	Restriction-Modification Activity			
	in vivo		in vitro	
	R	M	R	M
pUV2	n.d.	+	n.d.	n.d.
pURM4	+/-	+/-	+	+/-
pURM4ΔM	+/-	-	+	-
pURM4ΔR	-	+	-	+

FIG. 1. Restriction map and subclones of a 6.1-kb *Sst*I-*Xho*I fragment of plasmid pFV1 containing the *Mth*TI system. Arrows indicate the orientation of the ENase (R) and MTase (M) genes. The in vivo and in vitro ENase (R) and MTase (M) activities of the subclones are summarized. +/−, partial activity; n.d., not determined.

pUV2 carries the *Mth*TI R/M system (Fig. 1). For expression studies in *E. coli*, a 2.9-kb *Hin*dII-*Hin*dIII fragment from plasmid pUV2 was cloned into pUC18 digested with *Hin*dII and *Hin*dIII, yielding plasmid pURM4 (Fig. 1). Deletion of a 1.6-kb fragment between the *Pst*I site of the polylinker and the *Nsi*I site of the insert of plasmid pURM4 resulted in plasmid pURM4ΔM (Fig. 1). For construction of plasmid pURM4ΔR (Fig. 1), a 1.95-kb *Sca*I-*Pst*I fragment from plasmid pUV2 was cloned into pUC18 digested with *Sma*I and *Pst*I.

Enzymes and chemicals. Restriction enzymes, T4 DNA ligase, and Klenow DNA polymerase were obtained from Life Technologies Inc. (Gaithersburg, Md.). M.*Hae*III was purchased from New England Biolabs, Inc., Beverly, Mass. Oligonucleotides were synthesized with a Biosearch Cyclone DNA synthesizer (New Brunswick Scientific Corp.) at The Netherlands Institute for Dairy Research. [α-^{32}P]dATP (110 TBq/mmol) was purchased from Amersham International (Amersham, United Kingdom).

DNA manipulations and sequence analysis. Isolation of plasmids, transformation of *E. coli*, restriction mapping, agarose gel electrophoresis, and subcloning of DNA fragments were done essentially as described previously (31). For determination of the nucleotide sequence, DNA fragments from plasmid pUV2 were isolated from agarose gel, using a Geneclean kit as described by the supplier (Bio 101, La Jolla, Calif.), and subcloned into M13mp18/19 vectors. The nucleotide sequences of overlapping fragments of both strands were determined with use of vector- and insert-specific primers by the dideoxy-chain termination method (32) with the Sequenase kit according to the recommendations of the manufacturer (United States Biochemical Corp., Cleveland, Ohio). Computer analysis of sequence data was done with the University of Wisconsin Genetics Computer Group package (version 6.0) (7) and the CAOS/CAMM facilities at Nijmegen, The Netherlands.

Preparation of cell extracts. Ten-milliliter overnight cultures of *E. coli* HB101 carrying (part of) the *Mth*TI R/M system were diluted with 10 ml of medium and induced with 2 mM isopropylthiogalactopyranoside (IPTG). After 4 h of growth, cells were harvested by centrifugation and washed in prechilled lysis buffer (100 mM NaCl, 10 mM EDTA, 10 mM β-mercaptoethanol, 50 mM Tris-HCl [pH 7.6]). Finally, cells were resuspended in 0.05 volume of lysis buffer and disrupted by sonication. After heat treatment for 10 min at 65°C (34), the extract was centrifuged for 2 min at 15,000 × *g*. The supernatant was collected, adjusted to 0.4 μg of protein per μl, mixed with 1 volume of glycerol, and stored at −20°C.

For preparation of cell extracts from *M. thermoformicicum* THF, approximately 0.1 g of cells was resuspended in 0.5 ml of lysis buffer and disrupted by sonication as described above. After centrifugation, the supernatant, which contained about 3.5 to 5.0 μg of protein per μl, was collected, mixed with 1 volume of glycerol, and stored at −20°C.

Endonuclease assay. Standard assays with R.*Mth*TI were performed in a total volume of 10 μl containing 1 μg of DNA, 1 μl of restriction endonuclease buffer (1 M NaCl, 0.1 M $MgCl_2$, 0.01 M dithioerythreitol, 0.5 M Tris-HCl [pH 7.6]), and 1 μl of cell extract. After incubation for 1 h at 65°C, the resulting DNA fragments were analyzed by agarose gel electrophoresis.

Methylase assay. For methylase assays with M.*Mth*TI, 1 μl of cell extract was incubated with 1 μg of DNA in 50 mM NaCl–1 mM dithiothreitol–50 mM Tris-HCl (pH 8.5)–0.08 mM *S*-adenosylmethionine (AdoMet) in a total volume of 5 μl. After incubation for 1 h at 65°C, 12 μl of deionized H_2O, 2 μl of restriction endonuclease buffer, and 1 μl of extract prepared from *E. coli*(pURM4ΔM) or 10 U of R.*Hae*III were added, and the mixture was incubated for another hour at 65 or 37°C, respectively. Subsequently, the DNA was analyzed by agarose gel electrophoresis. Assays using M.*Hae*III were performed under similar conditions but at 37°C.

Primer extension. Approximately 1 μg of pUV1 DNA was incubated under endonuclease conditions with extract prepared from *E. coli* harboring plasmid pURM4ΔM or R.*Hae*III. Digestion of the DNA was checked by agarose gel electrophoresis. The DNA was deproteinized by a repeated double phenol-chloroform extraction and chloroform extraction, precipitated with ethanol, and finally resuspended in deionized H_2O. After alkali denaturation, the DNA was precipitated again and resuspended in 5 μl of deionized H_2O. Addition of 2 μl of buffer (0.1 M Tris-HCl [pH 8.0], 0.05 M $MgCl_2$) and 3 μl of primer (0.5 pmol/μl) was followed by annealing for 15 min at 37°C. The suspension was mixed with 1 μl of 0.1 M dithiothreitol, 2 μl of deoxynucleoside triphosphate (dNTP) solution (dCTP, dGTP, and dTTP; 1.5 μM each), 0.2 μl of [α-^{32}P]dATP (10 μCi/μl), and 5 U of Klenow DNA polymerase and incubated for 5 min at room temperature. After addition of 4 μl of chase solution (all four dNTPs; 0.5 mM each), the mixture was incubated for 30 min at room temperature; the reaction was stopped by addition of 8 μl of the stop solution used for the sequence reactions.

Nucleotide sequence accession number. The nucleotide sequence reported in this paper has been assigned the GenBank accession number M97222.

RESULTS

Identification and sequence analysis of the *Mth*TI R/M system. The thermophilic *M. thermoformicicum* strain THF harbors the 13.5-kb plasmid pFV1 (24). During restriction analysis of pFV1, we observed that its DNA was resistant to digestion with the GGCC-specific ENase R.*Hae*III. Upon cloning of the two *Xho*I-*Sst*I fragments of pFV1 in *E. coli* JM83 by using pUC19, it appeared that DNA of one of the resulting plasmids, pUV1, was sensitive to R.*Hae*III. In contrast, DNA of the other plasmid, pUV2 (Fig. 1), was only partially digested with R.*Hae*III despite the presence of multiple *Hae*III recognition sites in at least the vector part of pUV2. The degree of protection against digestion with R.*Hae*III increased when the *mcrB E. coli* strain HB101 was used to propagate pUV2 (Fig. 1). These results indicate that *M. thermoformicicum* THF contains an MTase activity, designated M.*Mth*TI, which modifies the R.*Hae*III recognition sequence and is probably encoded by the insert DNA of pUV2. Subcloning of pUV2 DNA followed by R.*Hae*III digestion studies allowed the location of the putative *mthTIM* gene on a 1.95-kb *Sca*I-*Pst*I fragment, contained in plasmids pURM4 and pURM4ΔR. These plasmids were partially and completely resistant to digestion by R.*Hae*III, respectively (Fig. 1; see below).

Nucleotide sequence analysis of the 2,879-bp *Hin*dII-*Xho*I insert present in pURM4 (Fig. 2) revealed two large, partially overlapping open reading frames (ORFs). The first ORF is located between positions 166 and 1032. This ORF contains an initiation codon (ATG) at position 187, which is preceded by a potential ribosome binding site (positions 171 to 181). If this initiation signal is used, the ORF could code for a protein with 281 amino acids (32,102 Da). Since plasmids containing this ORF resulted in an ENase activity in *E. coli* (Fig. 1; see below), we conclude that it represents the *mthTIR* gene. The second ORF ranges from positions 1026 to 2019 and is included in the 1.95-kb *Sca*I-*Pst*I fragment, which specifies MTase activity (see below). Therefore, we conclude that this ORF contains the *mthTIM* gene. A potential start site of the *mthTIM* gene is the second triplet of this ORF (ATG at position 1029), which partially overlaps with the opal stop codon (TGA) of the preceding *mthTIR* gene. The *mthTIM* gene is preceded by a putative ribosome binding site at positions 1014 to 1018. If this ATG is used, the *mthTIM* gene could code for a protein of 330 amino acids (37,360 Da).

Preceding the coding regions of each gene, two possible transcription initiation sites with similarities to a typical methanogenic promoter (4, 29) consisting of a consensus *boxA* and a *boxB* could be identified (Fig. 2), but their functionality has to be verified. Immediately following the *mthTIR* gene, a T stretch that could function as a transcription termination signal was observed (Fig. 2) (4). A second T-rich stretch was identified downstream of the translation termination codon of *mthTIM*. This putative transcription termination signal is contained within a 198-bp region consisting of several direct repeats (Fig. 2), which were not found elsewhere in the DNA of pFV1 (24).

Activity and recognition sequence of R.*Mth*TI. Plasmids pURM4 and pURM4ΔM, which both contain the *mthTIR* gene under control of the vector-located *lac* promoter (Fig. 1), appeared to encode active ENase, since *E. coli* HB101 harboring either of those plasmids showed low recovery and substantial degradation of plasmid DNA (data not shown). To analyze the in vitro ENase activity, cell extracts prepared from *E. coli* HB101 harboring pURM4ΔM were incubated with phage λ and pUV1 DNA at 65°C. The results showed that the number and size of the restriction fragments obtained with the extract were identical to those obtained by digestion with purified R.*Hae*III at 37°C (Fig. 3A). Similar digestion patterns were obtained with cell extracts prepared from cells of HB101 harboring pURM4 (Fig. 1) and *M. thermoformicicum* THF, although the activity of the latter extracts was much lower than that of the *E. coli* extracts. No ENase activity was found in control experiments using clones with either pURM4ΔR or pUC18 (data not shown). These results confirm that the *M. thermoformicicum* R.*Mth*TI activity is encoded by the cloned, pFV1-located *mthTIR* gene. In addition, the data indicate that the archaeal R.*Mth*TI recognizes the same DNA target sequence as does R.*Hae*III, i.e., GGCC (3).

To determine the exact position at which R.*Mth*TI cleaves the target sequence, primer extension experiments were performed, using R.*Mth*TI- or R.*Hae*III-digested pUV1 DNA as a template. The primer extension product of DNA incubated with R.*Mth*TI was the same size as that derived from incubation with R.*Hae*III (Fig. 4). Inspection of the sequence ladder run in parallel showed that both ENases cleave the target sequence to yield GG ↓ CC.

Activity of M.*Mth*TI in vivo and in vitro. To extend the observation that pUV2 DNA and DNA of some of its subclones (Fig. 1), when propagated in *E. coli* HB101, showed (partial) resistance to digestion with R.*Hae*III, those plasmids and control DNAs were incubated with R.*Mth*TI obtained from a cell extract of *E. coli* HB101 harboring pURM4ΔM. As was expected from the identical target site recognition of the two ENases, the same protection against digestion was observed. Remarkably, DNA of pUV2 and pURM4 was only partially protected against digestion by both R.*Hae*III and R.*Mth*TI, whereas DNA of pURM4ΔR was found to be fully protected against digestion by both ENases (not shown), suggesting different expression levels of the *mthTIM* gene in *E. coli* harboring these plasmids. Low M.*Mth*TI activity was also observed in extracts of *E. coli* HB101 harboring pURM4, which contains the complete *Mth*TI R/M system under control of the vector-located *lac* promoter (Fig. 1). Higher MTase activity was observed in cells containing pURM4ΔR, and almost complete protection of pUV1 DNA against restriction with either R.*Hae*III or pURM4ΔM-encoded R.*Mth*TI was found (Fig. 3B, lanes 1 to 4).

In vitro methylation assays using a cell extract of *E. coli* HB101(pURM4ΔR) clearly showed the dependence of the M.*Mth*TI activity on addition of the methyl group donor AdoMet (Fig. 3B, lanes 1 to 4). Similar dependence on AdoMet was obtained when pUV1 DNA was modified in vitro with M.*Hae*III at 37°C (Fig. 3, lanes 5 to 8). Furthermore, extracts from *M. thermoformicicum* THF also contained AdoMet-dependent MTase activity which gave protection against R.*Hae*III or R.*Mth*TI (not shown). However, this in vitro activity was lower than that observed in *E. coli* extracts harboring pURM4ΔR. These results demonstrate that the *mthTIR* and *mthTIM* genes are both components of the GGCC-recognizing R/M system *Mth*TI.

M.*Hae*III modifies the tetranucleotide target sequence at the third position, yielding GG^{me}CC (22). Since DNA modified by M.*Mth*TI is resistant to digestion with R.*Hae*III (Fig. 3B, lanes 2 and 4) and, vice versa, DNA modified by M.*Hae*III is also resistant to digestion of the pURM4ΔM-encoded R.*Mth*TI (Fig. 3B, lanes 5 and 7), it is likely that M.*Mth*TI modifies the target sequence at the same position as that modified by M.*Hae*III.

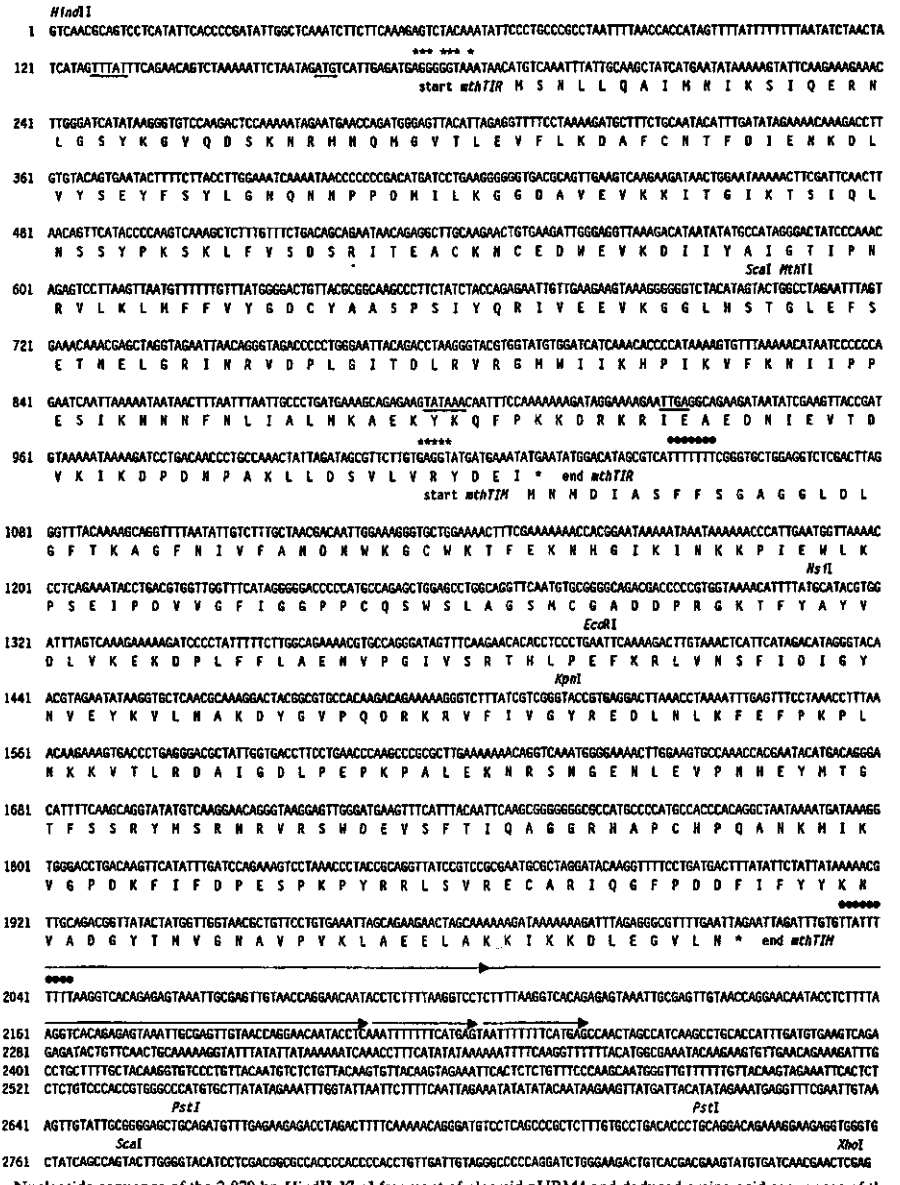


FIG. 2. Nucleotide sequence of the 2,879-bp *Hin*dII-*Xho*I fragment of plasmid pURM4 and deduced amino acid sequences of the *mthTIR* and *mthTIM* genes. The potential ribosome binding sites are marked by asterisks. Putative archaeal promoters and terminator structures (4, 29) are indicated by lines and circles, respectively. Arrows denote the location of direct repeats downstream from *mthTIM*.

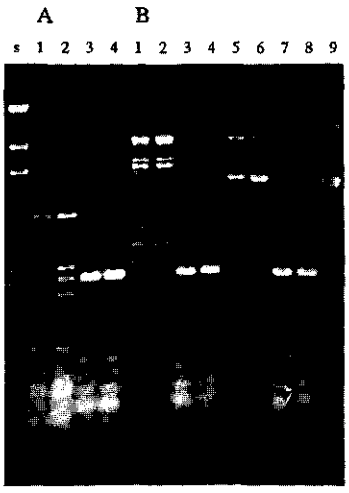


FIG. 3. In vitro ENase (A) and MTase (B) assays of extracts of *E. coli* HB101 containing cloned *mthTIR* and *mthTIM* genes and purified R.*Hae*III and M.*Hae*III. (A) Lanes: 1 and 3, DNA of phage λ and plasmid pUV1, respectively, incubated with extract prepared from *E. coli* harboring pURM4ΔM; 2 and 4, same DNA as in lanes 1 and 3 but incubated with R.*Hae*III (10 U). (B) Lanes: 1 to 4, plasmid pUV1 DNA incubated under MTase assay conditions with an extract prepared from *E. coli* harboring pURM4ΔR with (lanes 1 and 2) and without (lanes 3 and 4) AdoMet and reincubated under ENase assay conditions with extract prepared from *E. coli* harboring pURM4ΔM (lanes 1 and 3) and 10 U of R.*Hae*III (lanes 2 and 4); 5 to 8, same as lanes 1 to 4 except that M.*Hae*III (2.5 U) was used in the MTase assay; 9, undigested plasmid DNA from pUV1 (1 μg). Lane s, size marker phage λ DNA (0.5 μg) digested with *Hin*dIII.

Sequence comparison of R.*Mth*TI. The nucleotide and deduced amino acid sequences of the R.*Mth*TI enzyme were compared with published sequences of other GGCC-recognizing endonucleases R.*Ngo*PII (38) and R.*Bsu*RI (14). Interestingly, extensive sequence identity was found between R.*Mth*TI and R.*Ngo*PII (Fig. 5), and less, but still significant, identity was found between short stretches of amino acids of R.*Mth*TI and R.*Bsu*RI (data not shown). The high degree of identity (Fig. 5) between R.*Mth*TI and R.*Ngo*PII on the amino acid level (44.8% identity and 67.8% similarity)

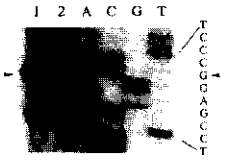


FIG. 4. Primer extension analysis of pUV1 DNA digested with R.*Hae*III (lane 1) or R.*Mth*TI (lane 2). The primer used had the sequence 5′-CTGTAGTTCAGGATC (positions 377 to 363 of plasmid pFV1) (24). Arrows indicate the extension product and its main 5′ nucleotide derived from incubation with both ENases. A shorter extension product attributed to nonspecific termination is also present. A DNA sequence ladder run in parallel is shown at the right.

FIG. 5. Amino acid sequence alignment of R.*Mth*TI and R.*NgoPII*. *Lines* denote identical amino acid residues, colons indicate conservative replacements, and single dots represent gaps. The alignment is based on the program BESTFIT of the University of Wisconsin Genetics Computer Group software (7). Regions with high amino acid identity are underlined.

and on the DNA level (53.3% identity) indicates a close relationship between the two enzymes, which is also reflected in their similar sizes (about 32 kDa each). Several highly conserved regions, comprising 13 to 34 amino acids each with 69 to 76% identity, could be identified (Fig. 5). Except for a conserved region located at the C terminus, both enzymes were most conserved in the central part.

Amino acid sequence comparison of M.*Mth*TI. Following alignments of amino acid sequences of several m^5C-MTases, a common architecture of these enzymes was postulated and 10 conserved blocks of amino acid residues were identified (19, 26). Those conserved domains are present with the same sequential order in the M.*Mth*TI protein sequence (Fig. 6). In addition, the variable region between boxes VIII and IX, which is believed to be responsible for sequence specificity (19), shared high similarity with the corresponding region found in the four characterized monospecific GGCC-recognizing MTases (M.*Hae*III, M.*Ngo*PII, M.*Bsu*RI, and M.*Bsp*RI). These four MTases can be subdivided by size: M.*Hae*III and M.*Ngo*PII (37.7 and 38.5 kDa, respectively) and M.*Bsu*RI and M.*Bsp*RI (49.6 and 48.3 kDa, respectively). M.*Mth*TI (37.4 kDa) showed the highest degree of amino acid similarity (69.2 and 68.3%, respectively) with M.*Hae*III and M.*Ngo*PII (53 and 51.9% identity, respectively; for comparison with M.*Ngo*PII, its alternative start site 12 amino acids downstream of the first Met was chosen). These high similarities are also reflected on the nucleotide level of the corresponding genes (55.3 and 55.9% identity, respectively). Comparison of M.*Mth*TI with M.*Bsu*RI and M.*Bsp*RI revealed 47.7 and 43.4% amino acid identity, respectively. This extensive sequence identity and the similar sizes of the enzymes suggest a common ancestor for the *mthTIM*, *haeIIIM*, and *ngoPIIM* genes.

Although the overall similarity between the MTases with GGCC specificity is very high, several residues within the conserved regions showed variations in the M.*Mth*TI enzyme. The most prominent one is a Pro residue at position 52

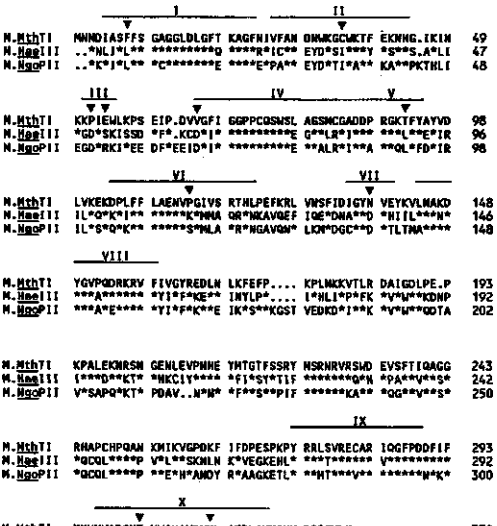


FIG. 6. Amino acid sequence alignment of M.*Mth*TI, M.*Hae*III, and M.*Ngo*PII, determined by using the program BESTFIT of the University of Wisconsin Genetics Computer Group software (7). The 11 putative N-terminal amino acid residues of the M.*Ngo*PII sequence have been omitted. The amino acids identical to those of M.*Mth*TI in sequences of M.*Hae*III and M.*Ngo*PII are indicated by asterisks. Single dots denote gaps. Regions marked I to X are conserved blocks found in the amino acid sequence of m^5C-MTases according to Posfai and coworkers (26). Positions marked with arrowheads represent unique amino acid substitutions in M.*Mth*TI compared with m^5C-MTases and 5′-GGCC-recognizing MTases.

(box III) of the M.*Mth*TI sequence, which is invariably an Asp in all m^5C-MTases analyzed to date (19, 26, 47). Similarly, other single amino acid substitutions at invariant positions of GGCC-recognizing MTases affect the local charge, e.g., replacement of the positively charged Arg residues by Pro and Thr residues at positions 114 (box VI) and 303 (box X), respectively, of the M.*Mth*TI enzyme. In general, there is a tendency toward more hydrophobic substitutions in the conserved regions in the M.*Mth*TI protein, although the overall hydrophobicity of M.*Mth*TI is similar to that of M.*Ngo*PII and M.*Hae*III.

DISCUSSION

In this report, we describe the identification and characterization of the GGCC-recognizing R/M system *Mth*TI from *M. thermoformicicum* THF, which to the best of our knowledge is the first published archaeal R/M system analyzed at the nucleotide level. Characterization of the *Mth*TI R/M system revealed new features with respect to its genomic location, similarity to other R/M systems, and thermostability. The *mthTIR* and *mthTIM* genes appeared to be located on the 13.5-kb plasmid pFV1 isolated from *M. thermoformicicum* THF (23). All plasmids isolated so far from methanogenic bacteria are cryptic (4). Therefore, pFV1 is the first plasmid in methanogens that codes for identified enzymatic activities.

It was previously reported that the *ngoPIIM* gene could be cloned only in an *mcrB E. coli* host (37). This limitation was attributed to the fact that the *ngoPIIM* gene encoded an MTase that generates methylated DNA with the sequence $GG^{me}CC$ which is a substrate for the *mcrB* endonuclease. Nevertheless, we could easily clone the *Mth*TI system in *E. coli* JM83, which is described as *mcrB* proficient (47). Since we observed that plasmids carrying the *mthTIM* gene were only partially resistant to R.*Hae*III when propagated in *E. coli* JM83, limited expression of the *mthTIM* gene could be a simple explanation for the viability of this host. Alternatively, it could be argued that M.*Mth*TI did not modify the target sequence to yield $GG^{me}CC$ but that either $GGC^{me}C$ or $GG^{hme}C^{hme}C$ (^{hme}C is hydroxymethylcytosine) was produced. However, these options are not very likely because R.*Hae*III is capable of cutting $GGC^{me}C$ (22) and DNA modified by M.*Mth*TI was resistant to cleavage by R.*Hae*III. In addition, hydroxymethylation of cytosine has not been described for type II MTases.

Plasmid pURM4ΔM, which contains only the *mthTIR* gene, could be maintained in a modification-deficient *E. coli* strain. Similar results have been described for the cloned *pstIR* and *taqIR* genes (10, 34). Since we found partial activity of R.*Mth*TI at 37°C, *E. coli* must contain efficient mechanisms to prevent cleavage by R.*Mth*TI or to repair cleaved DNA. This protection appears not to be complete, since only low amounts of pURM4ΔM DNA could be isolated and appeared to be partially degraded.

The thermophilic methanogens *M. thermoformicicum* and *M. thermoautotrophicum* are related species (40). It has been shown that *M. thermoautotrophicum* and other thermophilic bacteria contain low levels of m^5C (8). This has been explained by the fact that those organisms avoid production of m^5C because high temperatures favor deamination of m^5C ($m^5C \rightarrow T$), which would result in a point mutation (8). The presence of an m^5C-MTase in *M. thermoformicicum* THF suggests the presence of a mechanism either to avoid deamination of m^5C or to repair mismatched DNA.

Since *M. thermoformicicum* is a thermophilic organism, it was not unexpected to find that its *Mth*TI system showed activity at 65°C, the optimal growth temperature of the host. In fact, we found that R.*Mth*TI is highly active even at temperatures up to 75°C (unpublished results). All other R/M systems encoding an m^5C-MTase have been isolated from mesophilic organisms (45). The observation that R.*Mth*TI and M.*Mth*TI enzymes extracted from *E. coli* could be assayed at high temperatures in standard buffer systems indicates that the thermostability of the *Mth*TI enzymes is an intrinsic property. In the absence of a three-dimensional structure, comparison of primary sequences derived from homologous mesophilic and thermophilic proteins may give some indications as to the basis for thermostability. In general, thermostability of proteins can be attributed to an increased conformational stability. This may be realized by interactions between different parts of the protein, mediated by hydrophobic, electrostatic, or other interactions, that prevent denaturation at high temperatures (11). Notably, two kinds of differences which could contribute to such interactions were observed when the protein sequences of the *Mth*TI system were compared with those of their mesophilic counterparts (Fig. 5 and 6). Both R.*Mth*TI and M.*Mth*TI showed a shift of charged amino acid residues from the N-terminal to the C-terminal part of the protein

compared with the other enzymes, whereas the overall charge of the proteins remained the same. In addition, single amino acid substitutions in conserved regions of M.*Mth*TI tend to be less polar residues.

The M.*Mth*TI enzyme shares extensive amino acid similarity with the other GGCC-recognizing monospecific MTases M.*Ngo*PII and M.*Hae*III in conserved regions that are expected to be involved in substrate binding and recognition of the target sequence (19, 26) and methyl group transfer (6). This finding supports the conclusions drawn from analyses of several other R/M systems that the C-MTases seem to have evolved from a common origin (18). In contrast, ENases, even those with same sequence specificity, are usually different, indicating different evolutionary origins. The *Mth*TI system is an exception, since R.*Mth*TI showed a high degree of similarity to R.*Ngo*PII isolated from *N. gonorrhoeae*. R.*Mth*TI and R.*Ngo*PII are similar in size and are encoded by genes that have the same orientation with respect to the corresponding MTase gene. The same organization and homology to the *Ngo*PII system was reported for the isospecific R/M system *Fnu*DI from *Fusobacterium nucleatum* D, but unfortunately its sequence has not been published (45). In contrast, the R.*Hae*III protein sequence did not show similarities to that of R.*Ngo*PII and thus is probably also different from the R.*Mth*TI protein sequence, in accordance with the different organization of the *Hae*III R/M system (45).

The high similarity at the protein and DNA levels and the identical organizations of the *Mth*TI and *Ngo*PII R/M systems strongly suggest that the entire systems have been derived from a common ancestor and disseminated via horizontal gene transfer. Since *M. thermoformicicum* belongs to the domain *Archaea* and *N. gonorrhoeae* belongs to the domain *Bacteria*, this conclusion implies that gene transfer had occurred between these evolutionarily different groups of prokaryotes (46). If so, it is reasonable to assume that the direction of transfer is from the thermophilic *M. thermoformicicum* to the mesophilic *N. gonorrhoeae* rather than vice versa, since the *Mth*TI system is also active at low temperatures and would immediately confer a selective advantage when transferred to a mesophilic host. This is not to be expected when an R/M system from a mesophilic bacterium is introduced into a thermophilic organism. Among the possible avenues for gene transfer, transformation by DNA released into the environment is a likely one, especially since *N. gonorrhoeae* shows a natural competence for DNA uptake (5). Conjugation, however, cannot be excluded since it is known to play a significant role in gene transfer between distantly related organisms, including prokaryotes and eukaryotes (21). It is tempting to speculate that the plasmid location of the *Mth*TI system in *M. thermoformicicum* THF has been involved in the presumed transfer.

Comparison of the *Mth*TI and *Ngo*PII systems at the nucleotide level revealed that the high level of sequence identity did not extend beyond the coding regions of the ENase and MTase genes and could reflect the requirement for appropriate, host-specific expression signals, which are known to be very different between members of the *Archaea* and *Bacteria* (4).

ACKNOWLEDGMENT

We thank Rik Eggen for helpful discussions and advice on the manuscript.

REFERENCES

1. **Behrens, B., M. Noyer-Weidner, B. Pawlek, R. Lauster, T. S. Balganesh, and T. A. Trautner.** 1987. Organization of multispecific DNA methyltransferases encoded by temperate *Bacillus subtilis* phages. EMBO J. **6**:1137–1142.
2. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
3. **Bron, S., and K. Murray.** 1975. Restriction and modification in *B. subtilis*. Nucleotide sequence recognized by restriction endonuclease R.*Bsu*R from strain R. Mol. Gen. Genet. **143**:25–33.
4. **Brown, J. W., C. J. Daniels, and J. N. Reeve.** 1989. Gene structure, organization, and expression in archaebacteria. Crit. Rev. Microbiol. **16**:287–337.
5. **Cannon, J. G., and P. F. Sparling.** 1984. The genetics of the gonococcus. Annu. Rev. Microbiol. **38**:111–133.
6. **Chen, L., A. M. MacMillan, W. Chang, K. Ezaz-Nikpay, W. S. Lane, and G. L. Verdine.** 1991. Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase. Biochemistry **30**:11018–11025.
7. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
8. **Ehrlich, M., M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, and C. W. Gehrke.** 1985. DNA methylation in thermophilic bacteria: N^4-methylcytosine, 5-methylcytosine, and N^6-methyladenine. Nucleic Acids Res. **13**:1399–1412.
9. **Gibson, T. J.** 1984. Ph.D. thesis. University of Cambridge, Cambridge, United Kingdom.
10. **Gingeras, T. R., and J. E. Brooks.** 1983. Cloned restriction/modification system from *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA **80**:402–406.
11. **Jaenicke, R., and P. Zavodszky.** 1990. Proteins under extreme physical conditions. FEBS Lett. **268**:344–349.
12. **Kapfer, W., J. Walter, and T. A. Trautner.** 1991. Cloning, characterization and evolution of the *Bsu*FI restriction endonuclease gene from *Bacillus subtilis* and purification of the enzyme. Nucleic Acids Res. **19**:6457–6463.
13. **Kessler, C., and V. Manta.** 1990. Specificity of restriction endonucleases and DNA modification methyltransferases—a review. Gene **92**:1–248.
14. **Kiss, A., G. Posfai, C. C. Keller, P. Venetianer, and R. J. Roberts.** 1985. Nucleotide sequence of the *Bsu*RI restriction-modification system. Nucleic Acids Res. **13**:6403–6421.
15. **Klimasauskas, S., A. Timinskas, S. Menkevicius, D. Butkiene, V. Butkus, and A. Januliatis.** 1989. Sequence motifs characteristic of DNA [cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 methyltransferases. Nucleic Acids Res. **17**:9823–9832.
16. **Kruger, D. H., and T. A. Bickle.** 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol. Rev. **47**:345–360.
17. **Lange, C., A. Jugel, J. Walter, M. Noyer-Weidner, and T. A. Trautner.** 1991. Pseudo domains in phage-encoded DNA methyltransferases. Nature (London) **352**:645–648.
18. **Lauster, R.** 1989. Evolution of type II DNA methyltransferases: a gene duplication model. J. Mol. Biol. **206**:313–321.
19. **Lauster, R., T. A. Trautner, and M. Noyer-Weidner.** 1989. Cytosine-specific type II DNA methyltransferases: a conserved enzyme core with variable target-recognizing domains. J. Mol. Biol. **206**:305–312.
20. **Lunnen, K. D., R. D. Morgan, C. J. Timan, J. A. Krzycki, J. N. Reeve, and G. G. Wilson.** 1989. Characterization and cloning of *Mwo*I (GCN_7GC), a new type-II restriction-modification system from *Methanobacterium wolfei*. Gene **77**:11–19.
21. **Mazodier, P., and J. Davies.** 1991. Gene transfer between distantly related bacteria. Annu. Rev. Genet. **25**:147–171.
22. **Nelson, M., and M. McClelland.** 1989. Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res. **17**:r389–r415.
23. **Nölling, J., M. Frijlink, and W. M. de Vos.** 1991. Isolation and characterization of plasmids from different strains of *Methano-*

bacterium thermoformicicum. J. Gen. Microbiol. **137**:1981–1986.
24. **Nölling, J., F. J. M. van Eeden, and W. M. de Vos.** Submitted for publication.
25. **Patterson, N. H., and C. Pauling.** 1985. Evidence for two restriction-modification systems in *Halobacterium cutirubrum*. J. Bacteriol. **163**:783–784.
26. **Posfai, J., A. S. Bhagwat, G. Posfai, and R. J. Roberts.** 1989. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. **17**:2421–2435.
27. **Posfai, G., A. Kiss, S. Erdei, J. Posfai, and P. Venetianer.** 1983. Structure of the *Bacillus sphaericus* R modification methylase gene. J. Mol. Biol. **170**:597–610.
28. **Prangishvili, D. A., R. P. Vashakidze, M. G. Chelidze, and I. U. Gabriadze.** 1983. A restriction endonuclease *Sua*I from the thermophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett. **192**:57–60.
29. **Reiter, W. D., U. Hüdepohl, and W. Zillig.** 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. Proc. Natl. Acad. Sci. USA **87**:9509–9513.
30. **Roberts, R. J.** 1990. Restriction enzymes and their isoschizomers. Nucleic Acids Res. **18**:r2149–r2365.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Schmid, K., M. Thomm, A. Laminet, F. G. Lane, C. Kessler, K. O. Stetter, and R. Schmitt.** 1984. Three new restriction endonucleases *Mae*I, *Mae*II and *Mae*III from *Methanococcus aeolicus*. Nucleic Acids Res. **12**:2619–2628.
34. **Slatko, B. E., J. S. Benner, T. Jager-Quinton, L. S. Moran, T. G. Simcox, E. M. Van Cott, and G. G. Wilson.** 1987. Cloning, sequencing and expression of the *Taq*I restriction-modification system. Nucleic Acids Res. **12**:9039–9049.
35. **Slatko, B. E., R. Croft, L. S. Moran, and G. G. Wilson.** 1988. Cloning and analysis of the *Hae*III and *Hae*II methyltransferase genes. Gene **74**:45–50.
36. **Stephenson, F. H., B. T. Ballard, H. W. Boyer, J. M. Rosenberg, and P. J. Greene.** 1989. Comparison of the nucleotide and amino acid sequences of the *Rsr*I and *Eco*RI restriction endonucleases. Gene **85**:1–13.
37. **Sullivan, K. M., and J. R. Saunders.** 1988. Sequence analysis of the *Ngo*PII methyltransferase gene from *Neisseria gonorrhoeae* P9: homologies with other enzymes recognizing the sequence 5'GGCC-3'. Nucleic Acids Res. **16**:4369–4387.
38. **Sullivan, K. M., and J. R. Saunders.** 1989. Nucleotide sequence and genetic organization of the *Ngo*PII restriction-modification system from *Neisseria gonorrhoeae*. Mol. Gen. Genet. **216**:380–387.
39. **Thomm, M., G. Frey, B. J. Bolton, F. Laue, C. Kessler, and K. O. Stetter.** 1988. *Mvn*I: a restriction enzyme in the archaebacterium *Methanococcus vanielii*. FEMS Microbiol. Lett. **52**: 229–234.
40. **Touzel, J. P., E. Conway de Macario, J. Nölling, W. M. de Vos, T. Zhilina, and A. M. Lysenko.** 1992. DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol. **42**:408–411.
41. **Tran-Betcke, A., B. Behrens, M. Weidner-Noyer, and T. A. Trautner.** 1986. Chimeric multispecific DNA methyltransferases with a novel combinations of target recognition. Gene **42**:89–96.
42. **Walter, J., M. Noyer-Weidner, and T. A. Trautner.** 1990. The amino acid sequence of the CCGG recognizing DNA methyltransferase *M.Bsu*FI: implications for the analysis of the sequence recognition by cytosine DNA methyltransferases. EMBO J. **9**:1007–1013.
43. **Wilke, K., E. Rauhaut, M. Noyer-Weidner, R. Lauster, P. Pawlek, B. Behrens, and T. A. Trautner.** 1988. Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. EMBO J. **7**:2601–2609.
44. **Wilson, G. G.** 1991. Organization of restriction-modification systems. Nucleic Acids Res. **19**:2539–2566.
45. **Wilson, G. G., and N. E. Murray.** 1991. Restriction and modification systems. Annu. Rev. Genet. **25**:585–627.
46. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc. Natl. Acad. Sci. USA **87**:4576–4579.
47. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.

Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene

Jörk Nölling and Willem M. de Vos

Nucleic Acids Research **20**, 5047-5052.

Nucleic Acids Research, Vol. 20, No. 19 5047–5052

Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI from *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene

Jörk Nölling and Willem M.de Vos*
Bacterial Genetics Group, Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

Received July 14, 1992; Revised and Accepted September 9, 1992 EMBL accession no. X67212

ABSTRACT

Two CTAG-recognizing restriction and modification (R/M) systems, designated *Mth*ZI and *Mth*FI, were identified in the thermophilic archaeon *Methanobacterium thermoformicicum* strains Z-245 and FTF, respectively. Further analysis revealed that the methyltransferase (MTase) genes are plasmid-located in both strains. The plasmid pFZ1-encoded *mthZIM* gene of strain Z-245 was further characterized by subcloning and expression studies in *Escherichia coli* followed by nucleotide sequence analysis. The *mthZIM* gene is 1065 bp in size and may code for a protein of 355 amino acids (M, 42,476 Da). The deduced amino acid sequence of the M.*Mth*ZI enzyme shares substantial similarity with four distinct regions from several m^4C- and m^6A-MTases, and contains the TSPPY motif that is so far only found in m^4C-MTases. Partially overlapping with the *mthZIM* gene and in reverse orientation, an additional ORF was identified with a size of 606 bp potentially coding for a protein of 202 amino acids (M, 23.710 Da). This ORF is suggested to encode the corresponding endonuclease R.*Mth*ZI.

INTRODUCTION

Restriction and modification (R/M) systems are widespread among prokaryotes (1) and have been identified in about 25% of the presently examined species (2). Their major function is proposed to be the protection from contamination by DNA of foreign origin, particularly derived from bacteriophage infection (3). The most common R/M systems are those of the type II, consisting of two enzymatic activities, an endodeoxyribonuclease (ENase) and a DNA-methyltransferase (MTase), which both recognize an identical short DNA sequence (2). Genetic characterization of type II R/M systems have shown that the ENase and MTase genes most often are closely linked (2, 4).

Three types of MTases, forming different methylation products, can presently be distinguished: those generating 5-methylcytosine (m^5C), N4-methylcytosine (m^4C), and N6-methyladenine (m^6A) (2). Based on their amino acid similarities, those MTases can be grouped into two families, the m^5C- on the one and the m^4C- and m^6A-producing MTases on the other hand (5–9). In contrast to the MTases, ENases usually do not share significant similarities on protein sequence level, except for some isoschizomeric enzymes (10–12).

The thermophilic methanogen *Methanobacterium thermoformicicum* belongs to one of the three fundamentally different groups of living organisms, the domain *Archaea*, distinguished from the *Bacteria* and the *Eucarya* by various unique characteristics (13, 14). Several strains of the species *M.thermoformicicum* were found to harbor partially homologous plasmids (15). One of these plasmids, plasmid pFV1, carries the genes for the GGCC-specific *Mth*TI R/M system of *M.thermoformicicum* strain THF (12). In the present report, we describe the identification of two CTAG-recognizing R/M systems, designated *Mth*ZI and *Mth*FI, in *M.thermoformicicum* strains Z-245 and FTF, respectively, and the characterization of the plasmid-encoded *mthZIM* gene.

MATERIAL AND METHODS

Bacterial strains, vectors, and growth conditions

M.thermoformicicum strains Z-245 (DSM 3720), FTF (DSM 3012), THF (DSM 3848), SF-4 (K. Yamamoto, Osaka City University, Japan), FF1 (15), FF3 (15) and CSM3 (16) were cultivated on H_2/CO_2 as the sole carbon and energy source as described previously (15). *Escherichia coli* strains DH5α (Life Technologies) and HB101 (17) were used for subcloning and propagation, respectively, of derivatives of plasmids pUC18 and pUC19 (18). *E.coli* TG1 (19) served as host strain for phages M13mp18 and M13mp19 (18). The *E.coli* strains were routinely grown in L-broth (20) at 37°C. If appropriate, ampicillin (50 μg/ml), isopropyl-β-D-thiogalactopyranoside (1 mM), or 5-bromo-4-chloro-3-indolyl-β-D-galactoside (0.004%) was added to the culture medium.

* To whom correspondence should be addressed

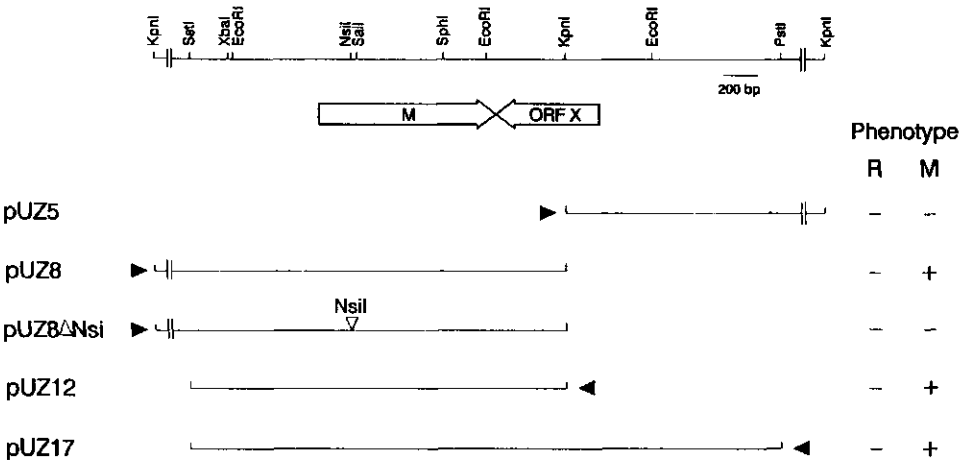


Figure 1. Restriction map and subclones of two adjacent *Kpn*I fragments (6.1 kb) of plasmid pFZ1. The orientation and location of the *mthZIM* gene (M) and ORF X are indicated. The *in vivo* ENase (R) and MTase (M) activity of the subclones are indicated on the right side; (▼) denote the location of the *lac*Z promoter of the vector part.

DNA manipulations and sequence analysis

Restriction endonuclease *Mae*I was used as recommended by the supplier (Boehringer Mannheim, Germany). All other restriction enzymes, T4 DNA ligase and Klenow DNA fragment were purchased from Life Technologies Inc. (Gaithersburg, Md.). DNA manipulations in *E.coli* were performed following established procedures (20). Total DNA was isolated from *M.thermoformicicum* as described previously (15). For nucleotide sequence determination, DNA fragments were isolated from agarose gels using a Geneclean kit as described by the supplier (Bio 101, La Jolla, Calif.) and subcloned into M13mp18/19. The nucleotide sequences of overlapping fragments of both strands were determined by the dideoxy chain termination method (21) with the Sequenase kit according to the recommendations of the supplier (United States Biochemical Corp. Cleveland, Ohio). [α-^{32}P]dATP (110 TBq/mmol) was purchased from Amersham (UK). Computer analysis of the sequence data was done with the GCG package version 6.0 (22) and the CAOS/CAMM facilities at Nijmegen, The Netherlands.

MthZI plasmid constructions

Plasmids pUZ5 and pUZ8 (Fig. 1) were derivatives of pUC19 which contain *Kpn*I-fragments of 2.1 kb and 3.7 kb, respectively, derived from plasmid pFZ1 from *M.thermoformicicum* strain Z-245 (15). Plasmid pUZ8ΔNsi (Fig. 1) was constructed by digestion of pUZ8 with *Nsi*I followed by treatment for 30 min with Klenow DNA polymerase in the presence of all four dNTPs (2 mM each) to remove 3'-overhanging ends. After ligation the DNA was used to transform *E.coli* HB101 cells. The resulting frameshift mutation in plasmid pUZ8ΔNsi was confirmed by restriction and sequence analysis (data not shown).

Several attempts failed to directly clone the 3.6 kb *Sst*I-*Pst*I fragment from plasmid pFZ1 (Fig. 1) into pUC18/19. Therefore, a plasmid that carries this fragment containing the *mthZIM* gene and open reading frame X (ORF X) was constructed on the basis of cloned pFZ1 DNA. For this purpose, pUZ5 and a derivative of pUZ8, which contained the 3.7 kb *Kpn*I-insert in opposite orientation relative to the pUC19-located *lac*Z-promoter, were used. For some manipulations, sites present in the polylinker of the vector parts were employed. First, a 2.3 kb *Sst*I-*Hind*III fragment of plasmid pUZ8 was cloned into pUC18 digested with *Sst*I and *Hind*III to yield plasmid pUZ12 (Figure 1). A short fragment of the polylinker located at the 5'-site of the truncated ORF X was removed from plasmid pUZ12 by digestion with *Sma*I and *Pst*I. Subsequently, a 1.2 kb *Sma*I-*Pst*I fragment containing the 5'-part of the ORF X gene was isolated from plasmid pUZ5 and inserted into the digested pUZ12 resulting in plasmid pUZ15. Due to this cloning strategy, which was chosen to avoid a possible lethal effect of transforming *E.coli* with a plasmid carrying an intact ORF X, plasmid pUZ15 harbors an ORF X that contains a frameshift mutation caused by a 14 bp *Kpn*I-*Sma*I-*Kpn*I polylinker fragment. Digestion of plasmid pUZ15 with *Kpn*I removed the 14 bp fragment and religation restored wild-type sequence in the pFZ1-derived insert DNA, resulting in plasmid pUZ17 (Figure 1). The integrity of the pFZ1 DNA at this location in pUZ17 was confirmed by sequence analysis (data not shown).

Preparation of cell-free extracts

Approximately 0.1 g frozen cells of *M.thermoformicicum* strain Z-245 or FTF were resuspended in 0.5 ml lysis buffer (0.1 M NaCl, 0.01 M EDTA, 0.01 M β-mercaptoethanol and 0.05 M Tris-HCl pH 7.6) and disrupted by sonication. After centrifugation for 2 min at 15000×g, the supernatant, which contained about 3.5 to 5.0 μg protein/μl, was collected, mixed with 1 vol glycerol and stored at −20°C before use. The extracts were found to be stable for at least 12 month without loss of the R/M activities.

Analysis of the *Mth*ZI and *Mth*FI ENase and MTase activities

The *in vitro* activities of the *Mth*ZI and *Mth*FI systems were assayed essentially as described previously (12) with extracts prepared from *M.thermoformicicum* strains Z-245 and FTF. The buffer used under ENase conditions contained 0.1 M NaCl, 0.2 M KCl, 0.01 M $MgCl_2$, 0.001 M dithioerythritol and 0.05 M Tris-HCl pH 7.6. As methyl-group donor 0.8 mM S-adenosylmethione (AdoMet) was used under MTase conditions. All assays were performed at 55°C. When specified, the ENase assays were deproteinized prior to agarose gel electrophoresis by incubation with proteinase K (0.1 mg/ml; Boehringer GmbH, Mannheim, Germany) in the presence of 0.5% sodium dodecyl sulfate for 15 min, followed by extraction with phenol/chloroform. Subsequently, the DNA was precipitated with ethanol and resuspended in ENase buffer. As substrate for the R/M enzymes, phage λ DNA and plasmid pUZ3, consisting of a 6.5 kb *Sma*I-fragment of plasmid pFZ1 from *M.thermoformicicum* strain Z-245 (15) cloned into pUC19, were used.

Primer extension

Primer extension was carried out as described previously (12). DNA from plasmid pUV1 that consists of a completely sequenced, 7.4 kb *Xho*I-*Sst*I fragment of plasmid pFV1 from *M.thermoformicicum* strain THF cloned into pUC19 (12, 16), was digested with R.*Mth*ZI or R.*Mae*I and used a template for extension of a primer (5'-CTGTAGTTCAGGATC) that binds to position 375 to 361 of the insert of pUV1.

RESULTS

Identification of the *Mth*ZI and *Mth*FI R/M systems

Following the characterization of the plasmid-encoded restriction-modification system *Mth*TI from *M.thermoformicicum* strain THF (12) we also analyzed other thermophilic members of the genus *Methanobacterium* for the presence of ENase activity (Table 1). Extracts prepared from *M.thermoformicicum* strains Z-245 and FTF, incubated with phage λ- or plasmid pUZ3 DNA under ENase conditions, were found to produce discrete DNA fragments indicating the presence of site-specific ENase activity (Figure 2A). Identical DNA patterns were derived from ENase activity of strains Z-245 and FTF, which was designated R.*Mth*ZI and R.*Mth*FI, respectively. In addition, *in vitro* methylation assays showed that the cell-free extracts of both strains exhibit AdoMet-dependent MTase activity that generated at least partially modified pUZ3 DNA (Figure 2A, lanes 9−12). Since this DNA methylation affects the ENase activities of strain Z-245 and FTF, it follows that the DNA is modified by the corresponding MTases M.*Mth*ZI and M.*Mth*FI, respectively. To verify the *in vivo* activity of both MTases, total DNA isolated from *M.thermoformicicum* strain Z-245 or FTF was incubated with extracts prepared from either strain. Figure 2B illustrates that DNA from strains Z-245 and FTF was resistant to digestion with R.*Mth*ZI (lanes 2, 3) but not with R.*Kpn*I which was used as a control (lanes 4 ,5). Similar results were observed using R.*Mth*FI (data not shown). The resistance of DNA from strain Z-245 and FTF to digestion with ENase from either strain confirmed the *in vivo* activity of M.*Mth*ZI and M.*Mth*FI. In contrast, DNA isolated from strain THF, which carries the GGCC-recognizing *Mth*TI system (12), was sensitive to R.*Mth*ZI (Figure 2B, lane 1). These results indicate that *M.thermoformicicum* strains Z-245 and FTF each harbor a R/M system, designated *Mth*ZI and *Mth*FI, respectively, with identical sequence specificity.

Table 1. Distribution of plasmid DNA and ENase activity in *M.thermoformicicum*

Strain	Plasmid (kb)	Recognition sequence of ENase activity	Reference
Z-245	pFZ1 (11.0)	CTAG	This work
FTF	pFZ2 (11.0)	CTAG	This work
THF	pFV1 (13.5)	GGCC	(12)
FF1, FF3, CSM3, SF-4	–	–	This work

Sequence specificity of R.*Mth*ZI and R.*Mth*FI

In order to determine the recognition sequence of R.*Mth*ZI and R.*Mth*FI we compared the restriction patterns produced by these ENases with those produced by several commercially available ENases. We identified the CTAG-recognizing ENase R.*Mae*I as an isoschizomer of R.*Mth*ZI and R.*Mth*FI since it generates restriction patterns identical to that produced by the *M.thermoformicicum* enzymes using phage λ- or plasmid pUZ3 DNA as substrate (Figure 2A, lanes 2−4 and 6−8). For some of the restriction fragments generated by digestion with extracts from the *M.thermoformicicum* strains, slightly aberrant mobilities were observed if samples were not deproteinized before gel electrophoresis, i.e. large fragments show an increased and small fragments a decreased mobility (see Figure 2A, lanes 1 and 5 for results with Z-245 extract). This most likely is due to components in the extracts with similar properties as the histone-like DNA binding protein HMt from *M.thermoautotrophicum* strain ΔH (23).

Since both R/M systems *Mth*ZI and *Mth*FI recognize the same target sequence, *Mth*ZI from strain Z-245 was used for a more detailed analysis. The exact cleavage position of R.*Mth*ZI was determined by primer extension experiments. The extension products of R.*Mth*ZI and R.*Mae*I digested DNA had identical sizes (Figure 3). Considering that the 5'-overhanging ends generated by both ENases were filled up by Klenow DNA polymerase and therefore the 3'-nucleotide of the extension products represent the 5'-nucleotide of the cleavage sites, comparison of the extension products with the parallel run sequence ladder revealed that both ENases cleave the target sequence to yield C↓TAG.

Identification of the *mthZIM* gene

M.thermoformicicum strain Z-245 and FTF harbor identical or nearly identical plasmids, pFZ1 and pFZ2, respectively, which are partially homologous to plasmid pFV1 isolated from strain THF (15). Since the R/M system *Mth*TI of strain THF is located on plasmid pFV1 (12) and no ENase activity was detected in extracts prepared from other *M.thermoformicicum* strains that lack plasmid DNA (Table 1), we anticipated that the R/M system *Mth*ZI of strain Z-245 could be located on plasmid pFZ1. To study this possibility, cloned plasmid pFZ1 DNA was tested for *in vivo* modification by incubation with Z-245 extract and subsequent analysis on agarose gel. All plasmids containing pFZ1 DNA were sensitive to R.*Mth*ZI except one, pUZ8, which was only partially digested with R.*Mth*ZI despite the presence of

multiple R.*Mth*ZI sites in at least the vector part of pUZ8. The protection against digestion with R.*Mth*ZI was even higher when *E.coli* HB101 was used to propagate plasmid pUZ8 (Figure 2B, lane 7). In addition, we tested pUZ8 DNA for sensitivity to R.*Xba*I since its target sequence TCTAGA includes the CTAG site recognized by R.*Mth*ZI. Sequence analysis (see below) showed that pUZ8 contains two R.*Xba*I restriction sites, one in the insert and one in the vector part. Plasmid pUZ8 was—in contrast to other plasmids tested—partially resistant to digestion with R.*Xba*I (Figure 2B, lane 11). In contrast, plasmid pUZ8ΔNsi, which contains a frameshift mutation, was found to be sensitive to digestion with both R.*Mth*ZI and R.*Xba*I (Figure 2B, lanes 8, 12). These results indicate that pUZ8 encodes a MTase activity which modifies the R.*Mth*ZI and the R.*Xba*I recognition sites. Subcloning of pUZ8 DNA followed by R.*Mth*ZI digestion allowed the location of the *mthZIM* gene on a 2.3 kb *Sst*I-*Kpn*I fragment, contained in plasmid pUZ12 (Figure 1). Although the *in vivo* modification of DNA containing the *mthZIM* gene clearly showed that this gene is functionally expressed in *E.coli*, no activity of M.*Mth*ZI was observed in *in vitro* assays performed as described earlier (12).

Sequence analysis (see below) revealed that the insert of pUZ12 contained, in close proximity to the *mthZIM* gene, part of an open reading frame (ORF X) which is a good candidate for the corresponding *mthZIR* gene. To exclude the possibility that the ENase gene is truncated in pUZ12 we constructed plasmid pUZ17 that contain the *mthZIM* gene and the entire ORF X (Figure 1). However, neither *in vitro* using cell-free extracts nor *in vivo* in restriction assays using bacteriophage λ, ENase activity was observed with *E.coli* HB101 harboring either plasmid pUZ17 or pUZ12 (data not shown).

Nucleotide sequence of the *mthZIM* gene

The complete nucleotide sequence of the 3561 bp *Sst*I-*Pst*I insert of plasmid pUZ17 has been determined. Analysis of the sequence revealed two large, converging and partially overlapping ORFs. The first ORF contains an ATG start codon at position 779 which is preceded by a potential ribosome binding site (RBS; position 770 through 775), and a TAG stop codon at position 1844. It could code for a polypeptide of 355 amino acids (M_r 42,476 Da). Since plasmids containing this ORF were resistant to digestion with R.*Mth*ZI but became sensitive to this ENase when a frameshift mutation was introduced (pUZ8ΔNsi), we conclude that it represents the *mthZIM* gene. The second ORF, designated ORF X, is located on the complementary strand and ranges from position 2488 to the termination codon TAA at position 1837.

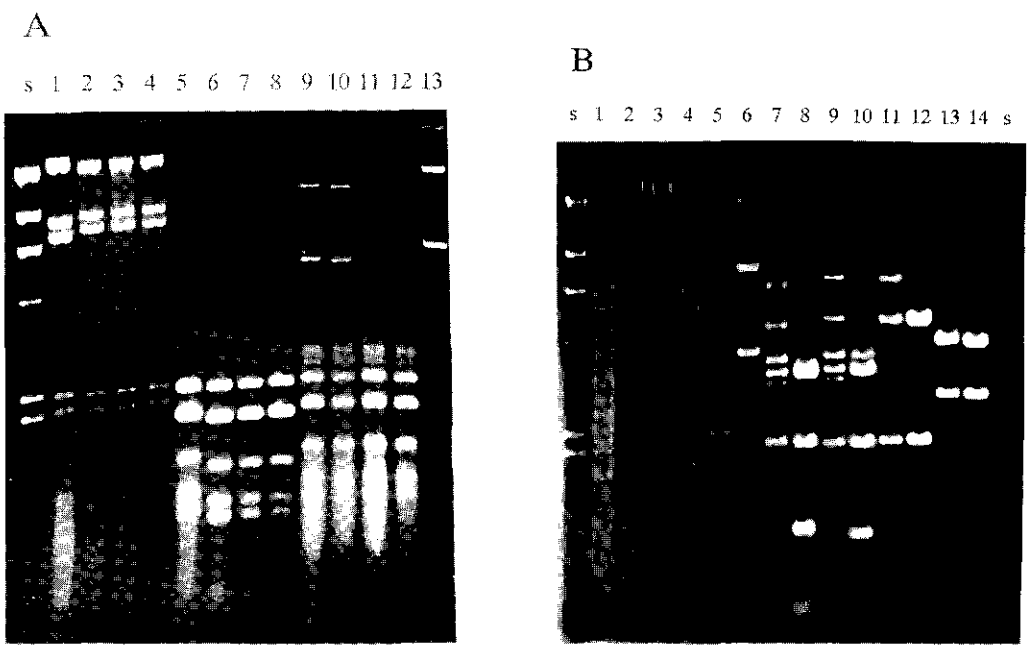


Figure 2. *In vitro* (A) and *in vivo* (B) activities of the *Mth*ZI and *Mth*FI R/M systems. (A) Lanes 1–8, ENase assays using extracts prepared from strain Z-245 (lanes 1, 5, not purified, lanes 2, 6, deproteinized), FTF (lanes 3, 7, deproteinized) and 5 Units R.*Mae*I (lanes 4, 8) with phage λ DNA (lanes 1–4) and pUZ3 DNA (lanes 5–8) as substrate; lanes 9–12, pUZ3 DNA incubated under MTase assay conditions with Z-245-extract (lanes 9, 11) and FTF-extract (lanes 10, 12), with (lanes 9, 10) and without AdoMet (lanes 11, 12) followed by digestion with R.*Mth*ZI (lanes 9, 11) and R.*Mth*FI (lanes 10, 12). Lane 13, undigested plasmid pUZ3 DNA (1 μg). Lane s, size marker phage λ DNA (0.7 μg) digested with R.*Hind*III. (B) Lanes 1–5, about 0.5 μg total DNA isolated from *M.thermoformicicum* strain THF (lane 1), Z-245 (lanes 2, 4) and FTF (lanes 3, 5) incubated under ENase conditions with extract prepared from strain Z-245 (lanes 1–3) and R.*Kpn*I (lanes 4, 5). Lanes 7–14, about 1.0 μg plasmid pUZ8 (lanes 7, 9, 11, 13) and pUZ8ΔNsi DNA (lanes 8, 10, 12, 14) incubated under ENase conditions with extract prepared from strain Z-245 (lanes 7, 8), R.*Mae*I (lanes 9, 10), R.*Xba*I (lanes 11, 12), and R.*Kpn*I (lanes 13, 14). Lane 6, undigested plasmid pUZ8 DNA (1.0 μg). Lane s, size marker phage λ DNA (0.7 μg) digested with R.*Hind*III.

A potential start codon (GTG at position 2443) is preceded by a sequence that may serve as RBS (position 2454 through 2449). If the GTG is used, the ORF X encodes a protein of 202 amino acids with a calculated molecular mass of 23,710 Da.

The regions flanking the *mthZIM* gene and ORF X were analyzed for archaeal transcription signals. Within an AT-rich 250 bp region preceding the *mthZIM* gene, that contains several direct repeats, various DNA stretches with similarities to the essential TATA box of archaeal promoters could be identified (24, 25). Inspection of the upstream region of the ORF X revealed no obvious promoter-like sequences. Downstream of *mthZIM* and ORF X, a conserved archaeal transcription termination signal consisting of a T-rich stretch (26) was identified.

Protein sequence comparisons

The deduced amino acid sequence of the M.*Mth*ZI protein was compared to that of known sequences of other MTases. No significant homology was observed with MTases that produce m^{5}C. M.*Mth*ZI shares, however, similarities with distinct regions from several m^{4}C- and m^{6}A-MTases. The MTases showing the highest similarity with M.*Mth*ZI include M.*Bam*H2 (27), M.*Bam*HI (9), M.*Sma*I (28) and M.*Cfi*9I (7), all of which were proven or assumed to be m^{4}C-MTases, and two representatives of the m^{6}A-MTase family, M.*Hinf*I (29) and M.*Rsr*I (30). Four blocks with conserved amino acid motifs, similar to those reported by Brooks et al. (1991), could be identified (Figure 4). The longest conserved segment, block D, contains the FXGXG-motif (motif DII in Figure 4; X = variable amino acid) common to almost all MTases (6–8). Block B contains the (S/D)PP(Y/F) motif which is present in all m^{6}A- and m^{4}C-MTases (7, 9, 31). Block A is less conserved in M.*Mth*ZI but is, similar to other m^{6}A- and m^{4}C-MTases, located at the N-terminus and contains the characteristic DXXE motif.

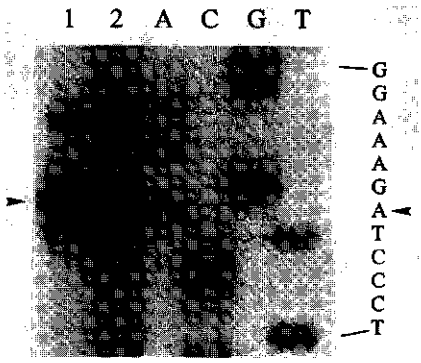


Figure 3. Primer extension analysis of the R.*Mth*ZI cleavage-site. Plasmid pUV1 DNA, digested with R.*Mae*I (lane 1) and R.*Mth*ZI (lane 2), respectively, was used as template for extension of the primer. Arrows indicate the major extension product and its corresponding nucleotide derived from incubation with both ENases. A shorter extension product attributed to aspecific termination is also present. A DNA sequence ladder run in parallel is shown at the right.

In addition to the conserved amino acid motifs identified by Brooks et al. (1991), the protein sequence comparison data revealed two other amino acid regions, which were found to be conserved in all m^{4}C-MTases (7, 27, 28, 32), except for M.*Mva*I (7), and in several m^{6}A-MTases such as M.*Hinf*I (29; Figure 4). Both regions are flanking the FXGXG motif (DII). One region, DI, contains the motif (P/S)XX(L/M)Y (Y = hydrophobic, non aromatic) and is located 22 amino acids N-terminal from the sequence FXGXG, whereas motif DIII, RXX(I/V)(G/S)XEX$_{11}$(R/K), is located 11 amino acids C-terminal from the FXGXG sequence. This latter motif contains three highly conserved charged residues.

Comparison of the amino acid sequence deduced from ORF X revealed no significant homology with other protein sequences found in the data base.

DISCUSSION

This report describes the identification of two novel R/M systems with the same sequence specificity, *Mth*ZI and *Mth*FI, from the archaeon *M.thermoformicicum*. In addition, the characterization of the first CTAG-recognizing methyltransferase, M.*Mth*ZI, is presented. Both *Mth*ZI and *Mth*FI R/M systems are produced by two plasmid-harboring strains, Z-245 and FTF, respectively. Further analysis revealed that the *mthZIM* gene is located on the 11 kb plasmid pFZ1 of strain Z-245 (15). Since the plasmids

```
                    A                                  B                                  C
M.MthZI   KNSADMNELKDKS  22   INLVV TSPPY PMVE  36   EVDRVTAPGGVVIINIGD  93
M.BamH2   IHNN*CVQFMKEN  18   *D*T* ***** DDLR  23   *LY***KE****VWVV**  54
M.BamHI   LYNG*CL**FKQV  133  VDTIF AD**F NNLD  151  *CI**LK***SLF*YNIP  193
M.SmaI    LDLF*QT*ECLE*  20   FDC** ***** WGLR  57   D*R*TLKDD*TLWL****  101
M.PvuII   MYIG*SL**LESF  42   *S**M ****F ALQR  60   V*NKKLK*D*SFVVDF*G  99
M.HinfI   ILKG*CI*KLKTI  25   *D*IF AD*** F*QT  43   *CK*ILKST*SIWVIGSF  98

                    DI                                 DII                                DIII
M.MthZI   PFELA YRLINMYSIMGDWVLDP FLGTG TTMIAAACAG RNSIGYELDHNFKDLIESR  I  275
M.BamH2   *EK** EDM*LSW*NE**I*F** *M*S* **AKM**LNN *KY**T*ISKEYC*IANE*  L  244
M.BamHI   SVK*L D*I*T*STNE**V**** *G*S* **FAVSEML* *KN**F**GNCEIIKERLK  N  361
M.SmaI    *R*M* RLCVLAG*RP*GK**** *F*S* **GVVCQELD *ECV*I**NEEYAS*AKE*  *  289
M.PvuII   *AK*P EFF*R*LTEPD*L*V*I *G*SN **GLV*ERES *KW*SF*KKPEYVAASAP*  F  308
M.HinfI   *ES*L *KV*LSS*KPN*V**** *F*** **GAV*KAL* **Y**I*REQKYI*VA*K*  L  255
```

Figure 4. Local amino acid sequence alignment of M.*Mth*ZI and other DNA methyltransferases. Amino acid sequences of four conserved regions (A–D) from M.*Mth*ZI, M.*Bam*H2, M.*Bam*HI, M.*Sma*I, M.*Pvu*II and M.*Hinf*I are shown. Region D is subdivided into three conserved amino acid blocks (DI-DIII). Residues identical to those in M.*Mth*ZI are denoted by asterisks. Amino acid motifs mentioned in the text are blocked. Numbers at the end of each line indicate the position of the last amino acid in the region.

pFZ1 and pFZ2 of the strains Z-245 and FTF are highly similar (15), it is very likely that the M.*Mth*FI enzyme is also encoded by plasmid pFZ2. Indeed, pFZ2 DNA showed strong hybridization with the *mthZIM* gene as probe (data not shown). In addition to the GGCC-specific, plasmid-located *Mth*TI system of strain THF (12), the *Mth*ZI and *Mth*FI systems represent another class of plasmid-encoded R/M systems in *M.thermoformicicum*. Since no ENase or MTase activity could be detected in other *M.thermoformicicum* isolates that appeared to be plasmid-free (Table 1), this would suggest that the occurrence of R/M systems in these methanogenic *Archaea* depends on the presence of extrachromosomal DNA elements.

Extracts prepared from *M.thermoformicicum* strains Z-245 and FTF contained high ENase and low MTase activities. The cloned *mthZIM* gene could be functionally expressed in *E.coli* as shown by *in vivo* modification of vector DNA carrying the complete *mthZIM* gene. No activity was, however, detected in *in vitro* MTase assays using extracts prepared from different *E.coli* strains with the *mthZIM* gene in either orientation with respect to the *lacZ* promoter of the vector. Since the *in vitro* MTase conditions used worked adequately with extracts prepared from the methanogenic hosts, a low expression level of the *mthZIM* gene or instability of the M.*Mth*ZI protein in *E.coli* could explain the inability to detect *in vitro* activity.

Comparison of the deduced amino acid sequence of the M.*Mth*ZI protein revealed four regions with significant similarity to m^4C- and m^6A-MTases but not to m^5C-MTases (Figure 4). Sequence comparisons did not reveal motifs that would allow a clear discrimination between m^4C- and m^6A-MTases. However, the deduced protein sequence of M.*Mth*ZI contains the TSPPY motif which so far has only been found in m^4C producing MTases (7, 28, 29, 32).

The recognition sequence CTAG of *Mth*ZI and *Mth*FI forms the core of the sequence TCTAGA, the target of the *Xba*I R/M system. We observed that DNA is protected to a similar extent against cleavage by either R.*Xba*I or R.*Mth*ZI when it is methylated by M.*Mth*ZI. If M.*Mth*ZI is a m^4C-MTase, it would modify the *Xba*I-site to yield $T^{m4}CTAGA$. Alternatively, if M.*Mth*ZI generates m^6A, it would modify the *Xba*I-site to yield $TCT^{m6}AGA$. Either modification pattern of the *Xba*I-site would be, in addition to the three already described (33, 34), a fourth one that prevent cleavage by R.*Xba*I.

Partially overlapping with the *mthZIM* gene and in opposite orientation, another ORF, ORF X, was identified. Although we have no direct evidence, ORF X could well represent the corresponding *mthZIR* gene. In favour of this possibility are the findings that (i) ENase and MTase genes most often are closely linked (2), (ii) the size of the protein encoded by ORF X is similar to that of other ENases, (iii) the *mthZIM* gene and ORF X constitute a 1.9 kb module in plasmid pFZ1 that is absent in plasmid pFV1, which in turn contains at this location a unique fragment encoding the *Mth*TI R/M system of strain THF (12, 16), and (iv) no ENase activity was observed in other, closely related strains of *M.thermoformicicum* that lack plasmid DNA (Table 1). The failure of detecting the expected ENase activity in *E.coli* might be ascribed to a low expression level of ORF X because of its unusual start codon GTG (35).

Among the large number of identified R/M systems (1) only four have previously been found to recognize the same sequence as *Mth*ZI and *Mth*FI (1, 2, 36, 37) indicating that CTAG is a rarely occurring target. Two of those systems, *Mae*I and *Mja*I, were identified in the methanogens *Methanococcus aeolicus* (36) and *Methanococcus jannaschii* (1). This suggest that, together with the here described *Mth*ZI and *Mth*FI, R/M systems recognizing CTAG are more common in *Archaea* than in *Bacteria*.

ACKNOWLEDGEMENT

We thank Rik Eggen for helpful discussions and advise on the manuscript.

REFERENCES

1. Roberts, R.J. (1990) Nucleic Acids Res. **18**, r2149–r2365.
2. Wilson, G.G. and Murray, N.E. (1991) Annu. Rev. Genet. **25**, 585–627.
3. Krüger, D.H. and Bickle, T.A. (1983) Microbiol. Rev. **47**, 345–360.
4. Wilson, G.G. (1991) Nucleic Acids Res. **19**, 2539–2566.
5. Lauster, R., Kriebardis, A. and Guschlbauer, W. (1987) FEBS Lett. **220**, 167–176.
6. Lauster, R. (1989) J. Mol. Biol. **206**, 313–321.
7. Klimasauskas, S., Timinskas, A., Menkevicus, S., Butkiene, D., Butkus, V. and Januliatis, A. (1989) Nucleic Acids Res. **17**, 9823–9832.
8. Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) Nucleic Acids Res. **17**, 2421–2435.
9. Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Moran, L.S., Slatko, B.E. and Benner,J.S. (1991) Nucleic Acids Res. **19**, 841–850.
10. Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) Gene **85**, 1–13.
11. Kapfer, W., Walter, J. and Trautner, T.A. (1991) Nucleic Acids Res. **19**, 6457–6463.
12. Nölling, J. and de Vos, W.M. (1992) J. Bacteriol. **174**, 5719–5726.
13. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA **287**, 4576–4579.
14. Zillig, W. (1991) Current Opinion in Genetics and Development **1**, 544–551.
15. Nölling, J., Frijlink, M. and de Vos, W.M. (1991) J. Gen. Microbiol. **137**, 1981–1986.
16. Nölling, J. (1992) Ph.D. thesis, Wageningen Agricultural University, The Netherlands (in preparation).
17. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. **41**, 459–472.
18. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
19. Gibson, T.J. (1984) Ph.D. thesis. University of Cambridge. Cambridge, United Kingdom.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
22. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395.
23. Musgrave, D.R., Sandman, K. and Reeve, J.N. (1991) Proc. Natl. Acad. Sci. USA **88**, 10397–10401.
24. Reiter, W.D., Hüdepohl, U. and Zillig, W. (1990) Proc. Natl. Acad. Sci. USA **87**, 9509–9513.
25. Hausner W., Frey, G. and Thomm, M. (1991) J. Mol. Biol. **222**, 495–508.
26. Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) CRC Crit. Rev. Microbiol. **16**, 287–337.
27. Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. (1990) Nucleic Acids Res. **18**, 4002.
28. Heidmann, S., Seifert, W., Kessler, C. and Domdey, H. (1989) Nucleic Acids Res. **17**, 9783–9796.
29. Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) Gene **70**, 387–392.
30. Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumport, R.I. (1989) Nucleic Acids Res. **17**, 10403–10425.
31. Chatterjee, D.K., Hammond, A.W., Blakesley, R.W., Adams, S.M. and Gerard, G.F. (1991) Nucleic Acids Res. **19**, 6505–6509.
32. Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M. and Blumenthal, R.M. (1989) Nucleic Acids Res. **17**, 4161–4175.
33. Van Cott, E.M. and Wilson, G.G. (1988) Gene **74**, 55–59.
34. McClelland, M. and Nelson, M. (1992) Nucleic Acids Res. **20**, 2145–2157.
35. Gold, L. and Stormo, G. (1987). In: *Escherichia coli* and *Salmonella typhimurium*; Neidhardt, C.(ed); American Society for Microbiology, Washington D.C.; Vol.2, 1302–1307.
36. Schmid, K., Thomm, M., Laminet, A., Lane, F.G., Kessler, C., Stetter, K.O. and Schmitt, R. (1984) Nucleic Acids Res. **12**, 2619–2628.
37. New England Biolabs, Inc. (1992) Product catalogue (2/92).

Summary and concluding remarks

The identification of the *Archaea* as a third primary lineage of life and their adaptation to extreme environmental conditions have generated considerable interest in the molecular biology of these organisms. Most progress in the investigation of archaeal mobile genetic elements, i.e. viruses, plasmids and insertion sequences, has been made in the halophilic branch while only limited knowledge was available about mobile elements of methanogens (reviewed in Chapter 1). The aim of this thesis was therefore to get more insight into the molecular genetics of mobile elements from methanogens. Thermophilic species of the genus *Methanobacterium* were used as model organisms in this study, since they are among the best characterized methanogens.

Phylogenetic studies of the species *M.thermoformicicum* and strains of *M.thermoautotrophicum* are described in Chapter 2. The comparison of two variable 16S rRNA regions allowed the conclusions that (i) *M.thermoformicicum* consists of two groups of strains, the Z-245-group (including strains Z-245, FTF, THF, FF1, FF3 and CSM3) and the CB12-group (consisting of the strains CB12, SF-4, and HN4), which probably constitute different species, (ii) *M.thermoautotrophicum* ΔH is closely related to the Z-245-group, and (iii) *M.thermoautotrophicum* Marburg belongs to neither group and most likely represents a different species (Table 1). This classification of *M.thermoformicicum* and *M.thermoautotrophicum* strains obtained by comparative 16S rRNA analysis is in line with recently reported results derived from DNA-DNA hybridization studies (8). Both approaches are based on genotypic characters which generally are more reliable for the determination of phylogenetic relationships than classical phenotypic ones. Therefore, the proposed reclassification of the examined thermophilic *Methanobacterium* strains may reflect their phylogeny more accurately than the current classification. Moreover, some other characters of the examined methanogens including genetic fingerprints with *fdh*A (Chapter 2 and 4), the sensitivity to phage ΦF1 (Chapter 3) and the distribution of FR-I (Chapter 6) provide additional evidence for the proposed new taxonomy of *M.thermoformicicum* and *M.thermoautotrophicum* (Table 1). Another phylogenetically relevant aspect reported in Chapter 2 was the finding that the non-formate utilizers *M.thermoautotrophicum* ΔH and Marburg contain sequences similar to the *fdh*A and *fdh*B genes from *M.formicicum*. The presence of *fdh*AB-like sequences in strain ΔH and Marburg suggest that both *M.thermoformicicum* and *M.thermoautotrophicum* are descendants of a common formate-utilizing ancestor.

It has recently been shown that *M.thermoautotrophicum* ΔH and Marburg differ clearly in their genome size (1.725 versus 1.623 kb, respectively) and the position of restriction sites in the chromosomal DNA for the endonucleases *Not*I, *Pme*I and *Nhe*I (7). These findings are likely to reflect the limited homology of both *M.thermoautotrophicum* strains on the genomic level. Similarly, also strain ΔH and *M.thermoformicicum* THF have been found to differ considerably in genome size (1.725 versus 1.600 kb, respectively) and *Not*I restriction pattern (7). This would contradict the results obtained by DNA-DNA similarity studies and comparative 16S rRNA sequence analysis which indicate a close relationship of these two strains (Chapter 2). However, the results of the genomic analysis of strain THF have to be

interpreted with caution for two reasons. Firstly, the *Not*I-site (5'-GCGGCCGC) comprises the target site of the GGCC-recognizing restriction-modification system *Mth*TI harbored by strain THF (see Chapter 7). As the consequence, the methyltransferase of the *Mth*TI system probably modifies the *Not*I-sites contained in DNA from strain THF to yield 5'-GCGG^{me}CCGC which is resistant against cleavage by *Not*I (4). The reported two *Not*I restriction fragments of 1.350 and 250 kb generated from THF DNA (7) may therefore not reflect the real number of *Not*I-sites in the chromosome of strain THF. Secondly, the enormous size of the large *Not*I fragment probably does not allow an accurate determination of its size.

Table 1. Classification of thermophilic *Methanobacterium* strains

Organism		Plasmid (kb)	Homology with FR-I	Homology with FR-II	Phage sensitivity ΦF1	Phage sensitivity ΦF3	Homology[a] group
M.thermoformicicum	CB12	—	—	—	—	—	CB12
	SF-4	—	—	—	—	—	CB12
	HN4	—	—	+	—	—	CB12
	Z-245	pFZ1 (11.0)	+	—	+	—	Z-245
	FTF	pFZ2 (11.0)	+	—	+	—	Z-245
	THF	pFV1 (13.5)	+	+	—	—	Z-245
	FF1	—	+	—	+	—	Z-245
	FF3	—	+	—	+	+	Z-245
	CSM3	—	+	+	+	—	Z-245
M.thermoautotrophicum	ΔH	—	+	—	+	—	Z-245[b]
	Marburg	pME2001 (4.4)	—	—	—	—	Marburg[b]

[a] Based on DNA-DNA similarity, 16S rRNA signature and genetic fingerprint with *fdh*A.
[b] Unique hybridization pattern obtained with *fdh*A.

Chapter 3 describes the characterization of the novel archaeal phages ΦF1 and ΦF3 that are able to infect several thermophilic strains of *Methanobacterium* (Table 1). Both phages differ with respect to their host range and the topology of their double-stranded DNA genomes. While phage ΦF1 has a broad host range and contains a linear, approximately 85-kb genome, ΦF3 was specific for *M.thermoformicicum* FF3 and contained a circular, approximately 36-kb genome. No similarity was found between the genomes of ΦF1 and ΦF3 nor between both phages and genomic DNA from different *Methanobacterium* strains or from phage ψM1 of *M.thermoautotrophicum* Marburg.

The isolation and initial molecular analysis of three plasmids, pFV1 (13.5 kb), pFZ1 (11.0 kb) and pFZ2 (11.0 kb), harbored by *M.thermoformicicum* strains THF, Z-245 and FTF, respectively, is reported in Chapter 4. Using cloned pFZ1 DNA as hybridization probe, the plasmids were found to constitute a family of highly related elements. Whereas pFZ1 and pFZ2 are probably identical, only partial similarity was observed between pFZ1 and pFV1. Furthermore, genetic fingerprinting experiments using *fdh*A from *M.formicicum* as hybridization probe allowed a subdivision of the species *M.thermoformicicum* into the CB12-group and the Z-245-group, comprising among others the plasmid-harboring strains Z-245, FTF and THF (Table 1).

The relatedness of the plasmid DNA was confirmed by sequence analysis of pFV1 and pFZ1 (Chapter 5). Comparison of the primary sequence of pFV1 (13513 bp; see also Appendix A) and pFZ1 (11014 bp; see also Appendix B) revealed a modular organization of the plasmid genomes: a backbone structure, conserved on the sequence level and in overall gene order, which is interspersed with plasmid-specific elements. The organization of the *M.thermoformicicum* plasmids resembles that of prokaryotic chromosomes both of which are subjected to DNA rearrangements, in particular insertions and/or deletions, thereby retaining a basic genome organization. The high sequence similarity and the comparable genetic maps suggest that pFV1 and pFZ1 (and pFZ2) have originated from a common ancestor. If so, the plasmids probably share homologous functions necessary for plasmid replication and maintenance which should be located within regions with high interplasmid similarity, i.e. the plasmid backbone. Those essential functions may include two large palindromic regions and a putative gene encoding a NTP-binding protein which are contained within the conserved backbone structure. In contrast, the plasmid-specific sequence blocks probably represent accessory elements which do not specify essential plasmid functions.

Both plasmids pFV1 and pFZ1 harbor sequences with similarity to chromosomal DNA of different thermophilic methanogens (Chapter 6). One of those, named FR-I, represents an accessory element of pFV1 with chromosomal counterparts in several *M.thermoformicicum* strains and *M.thermoautotrophicum* ΔH (Table 1). Comparison of the plasmid-derived and chromosomal FR-I elements revealed that FR-I has a size of 1.5 kb and exists in variants which differ in the organization of two subfragments. Remarkably, the corresponding subfragments of either FR-I element were identical on the nucleotide sequence level. Although FR-I lacks terminal inverted repeats, the presence of terminal direct repeats and its occurrence in multiple copies suggest that FR-I represents a new type of archaeal insertion sequences. A second element, termed FR-II, is part of both plasmids pFV1 and pFZ1, and the chromosome of *M.thermoformicicum* THF, CSM3 and HN4. Sequence analysis of the two plasmid- and one chromosome-derived FR-II elements showed that they are highly similar and may code for a protein with yet unknown function. In contrast to FR-I, FR-II is present in a single chromosomal copy and does not contain terminal repeats. Its presence in plasmid- and chromosomal DNA suggests, however, that FR-II is mobile or has been mobilized.

Each plasmid contains an accessory element encoding components of a type II R/M system: the GGCC-recognizing system *Mth*TI carried by pFV1 from strain THF (Chapter 7) and the CTAG-recognizing systems *Mth*ZI and *Mth*FI located on pFZ1 and pFZ2 from strain Z-245 and FTF, respectively (Chapter 8). These findings demonstrated for the first time plasmid-encoded enzymatic activities for methanogens. The R/M systems *Mth*TI and *Mth*ZI have been characterized in detail by their cloning and expression in *Escherichia coli*. Strikingly, both R/M modules are located within the same plasmid backbone region suggesting that this part of the plasmids has undergone substantial genomic rearrangements: either deletion or insertion of a single R/M module from an ancestral plasmid that contained both or none of the R/M cassettes. The presence of R/M cassettes would support the concept of a modular evolution, i.e., the mobilization of functional units within the same or between different replicons. Characterization of methanogenic R/M systems with the same specificity as the plasmid-located *Mth*TI and *Mth*ZI systems may provide insight in the relationships of archaeal R/M systems and possible mechanisms of dissemination of R/M modules.

Supporting evidence for a modular evolution is provided by the similarity between the two GGCC-recognizing R/M systems *Mth*TI from *M.thermoformicicum* THF and *Ngo*PII from *Neisseria gonorrhoeae* (Chapter 7). In contrast to other R/M systems with GGCC specificity,

*Mth*TI and *Ngo*PII comprise endonuclease- and methyltransferase genes which both were similar on the nucleotide and the deduced amino acid sequence level and exhibit an identical organization (Figure 1). These findings suggest that the *Mth*TI and the *Ngo*PII systems are homologous and have been disseminated via horizontal gene transfer. If so, gene transmission would have occurred between members of the evolutionary different domains *Archaea* and *Bacteria*.

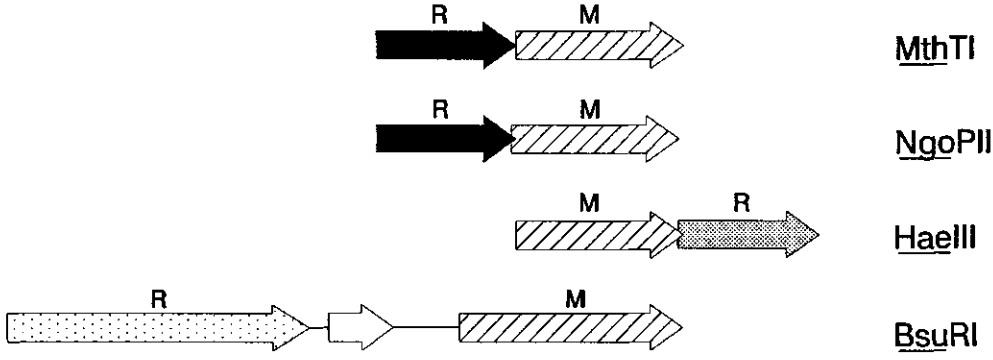


Figure 1. Gene organization of GGCC-recognizing restriction-modification systems. Homologous endonuclease (R) and methyltransferase (M) genes are indicated by similar shading.

An interesting property of the *Mth*TI system is the fact that it is present in a thermophilic organism. In contrast, all other known GGCC-recognizing, or, more general, m^5C-producing R/M systems, have been isolated from mesophilic organisms. The absence of such R/M systems in thermophiles has been attributed to an accelerated deamination rate of m^5C ($m^5C \rightarrow T$) at high temperatures generating G-T mismatches in double-stranded DNA which, if not corrected, would result in point mutations (1) (Figure 2). The thermophilic *M.thermoformicicum* THF apparently contains a mechanism which either avoids deamination of m^5C or is capable of G-T mismatch correction. The latter possibility seems to be realized in strain THF since the pFV1-located ORF10 has the capacity to code for a protein with significant similarity to *E.coli* DNA mismatch repair enzymes (Chapter 5; Figure 2). Similar to the organization of functionally related genes into operons, ORF10, *mthTIM* and *mthTIR* are located on the same pFV1 module (Figure 2) and may form a functional unit that specifies components of a restriction-modification-repair (RMR) system. Additionally, a putative fourth gene, ORF9, is located on the RMR module (Figure 2). Although no function could be assigned to the deduced ORF9 product, a participation in the activity of the RMR system would be plausible. Besides a possible common regulation, the observed gene clustering may also be necessary to ensure survival of a potential thermophilic recipient which acquires the RMR module via lateral gene transfer. However, if the recipient is a mesophilic organism, the

DNA repair functions would probably be dispensable and may be lost after transfer. Assuming that a transfer of the entire RMR module has occurred from the thermophilic *M.thermoformicicum* THF to the mesophilic *N.gonorrhoeae*, this may explain why the similarity included the genes of the *Mth*TI and *Ngo*PII systems but did not extend beyond adjacent sequence regions.

The major function of R/M systems is proposed to be the protection of the resident DNA from contamination by for instance phage DNA (Figure 2). The same function may be supposed for the discussed R/M systems from *M.thermoformicicum*. However, one would expect that R/M systems which generate thermostable modification products such as m^4C and m^6A would be more advantageous for a thermophilic host than those that produce m^5C, since deamination of the latter methylation product increases the chance on point mutations which have to be compensated by mismatch repair. Nevertheless, the m^5C-generating *Mth*TI R/M system obviously proved to be successful under thermophilic conditions since it is maintained by *M.thermoformicicum* THF. A possible explanation for the presence of the *Mth*TI system in strain THF may be that the system provides another advantage to its host, in addition to restriction of incoming phage DNA by the *Mth*TI endonuclease. As discussed in Chapter 5, deamination of m^5C may serve as additional protection mechanism against phages that contained m^5C-methylated genomes since, in contrast to the host, these phages are not be able to correct mismatches.

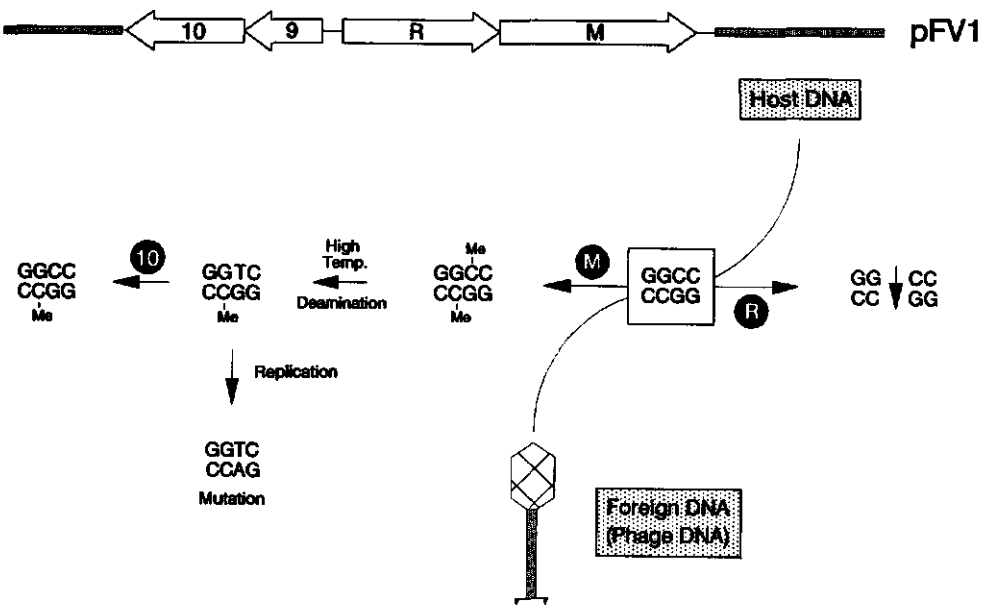


Figure 2. Schematic illustration of the proposed function of the restriction-modification-repair system harbored by plasmid pFV1 from *M.thermoformicicum* THF.

Analysis of the abundance of the tetranucleotide sequences GGCC and CTAG, the targets of the *Mth*TI and *Mth*ZI R/M systems, in DNA of thermophilic *Methanobacterium* strains revealed some interesting results. The sequence data used for analysis comprised 48.4 kb of chromosomal and plasmid DNA, the majority of which were obtained from the EMBL and GenBank data bases (Table 2). The tetranucleotide GGCC was found 228 times within 48.4 kb, which statistically corresponds with 1 tetranucleotide per 212 bp. This value is almost equal to the expected random-frequency for a tetranucleotide which is 1 per 256 bp. In contrast, only 13 CTAG tetranucleotides were detected, corresponding with 1 per approximately 3700 bp. A similar low abundance of CTAG (1 per 2700 bp) has been reported for the *E.coli* genome (5). This finding clearly shows that the recognition sequence of the *Mth*ZI system, in contrast to that of the *Mth*TI system, is rare in DNA from thermophilic *Methanobacterium* strains. However, a considerable different result was obtained from analysis of the plasmid DNA harbored by *M.thermoformicicum*. Compared to the values observed for the genomic DNA from thermophilic members of *Methanobacterium*, plasmid pFV1 exhibits a lower frequency of GGCC (1 site per 610 bp; 22 GGCC-sites per pFV1 genome) but shows a significantly higher abundance of CTAG (1 per 670 bp; 20 CTAG-sites per pFV1 genome)

Table 2. Distribution of GGCC- and CTAG-sites in DNA of thermophilic *Methanobacterium* strains

Gene or element	Organism	Size (kb)	Number of tetranucleotides		
			Expected[b]	Observed	
				GGCC	CTAG
sod	*M.thermoautotrophicum* sp.	1.4	5	5	0
purE	*M.thermoautotrophicum* sp.	1.4	5	4	0
rpsTUVX	*M.thermoautotrophicum* Winter	7.5	29	50	1
hmtA	*M.thermoautotrophicum* ΔH	1.1	4	4	1
mvhGDGA	*M.thermoautotrophicum* ΔH	5.3	21	20	1
frhADGB	*M.thermoautotrophicum* ΔH	3.9	15	19	5
trpABCDGE	*M.thermoautotrophicum* Marburg	7.4	29	40	1
ileS	*M.thermoautotrophicum* Marburg	4.2	16	28	1
mcrBGA	*M.thermoautotrophicum* Marburg	6.2	24	15	3
pME2001[c]	*M.thermoautotrophicum* Marburg	4.4	17	18	0
FR-I	*M.thermoformicicum* CSM3	3.1	12	12	0
FR-II	*M.thermoformicicum* CSM3	2.7	11	13	0
	Total	48.4	189	228	13
pFV1[c]	*M.thermoformicicum* THF	13.5	53	22	20
pFZ1[c]	*M.thermoformicicum* Z-245	11.0	43	28	6

[a] Comprising coding regions and flanking sequences.
[b] Random-frequency (on average 1 site per 256 bp).
[c] Complete nucleotide sequence

(Table 2). While pFZ1 contains an comparable number of GGCC-sites (28 per plasmid genome with a frequency of 1 per 390 bp), the CTAG-frequency is substantially lower in pFZ1 from strain Z-245 (1 per approximately 1800 bp; 6 per pFZ1 genome) (Table 2). These results are interpreted as follows:

(i) The low abundance of CTAG in the DNA of thermophilic *Methanobacterium* strains and *E.coli* suggests that a similar mechanism is responsible for this bias. It has recently been reported that the underrepresentation of several tetranucleotides, including CTAG, in the genome of *E.coli* may be the result of selection against target sequences of the *vsr* gene product (3), an enzyme involved in G-T to G-C mismatch correction (2). Usually, those target sequences of the Vsr endonuclease are generated by deamination of m^5C in the sequence $C^{me}C(A/T)GG$, the product of the *dcm* MTase. However, the *vsr* gene product displays a relative relaxed specificity since it also recognizes G-T mismatches in the sequence C<u>T</u>(A/T)G or <u>T</u>(A/T)GG where the underlined T is mismatched with G, and it does not require methylation of the unmutated strand (2). Consequently, T-G mismatches produced during recombination or DNA replication would, without determination which of both nucleotides is the incorrect one, sequence-specifically be corrected to C-G pairs. The tetranucleotide C<u>T</u>AG is one of the possible target sequences of the Vsr endonuclease and would be 'repaired' to CCAG, resulting in an elimination of CTAG-sites in time. A similar mechanism may be responsible for the low CTAG abundance in DNA of thermophilic members of *Methanobacterium*.

(ii) Compared to the genomic DNA of thermophilic *Methanobacterium* strains, plasmid pFV1 of strain THF showed a five-fold increased number of CTAG tetranucleotides. This significant difference may indicate that selection against CTAG is absent in strain THF. Alternatively, pFV1 (and the related pFZ1 and pFZ2) does not originate from DNA of a thermophilic *Methanobacterium* strain. The relatively low GC content of pFV1 (41.8%) compared to that of strain THF (49.6%) (8) would support the latter possibility.

(iii) Although pFZ1 of strain Z-245 is highly related to pFV1, the abundance of the tetranucleotide CTAG is clearly lower in pFZ1. Comparison of the nucleotide sequence of plasmid pFZ1 with the one from pFV1 showed that in particular CTAG-sites located within putative coding regions were absent in pFZ1. The lower frequency of CTAG-sites in pFZ1 correlates with the presence of the CTAG-recognizing *Mth*ZI system that is encoded by pFZ1. A possible explanation would be that the MTase M.*Mth*ZI provides only an incomplete protection against cleavage of the corresponding ENase which would result in selection against CTAG-sites. Another possibility may be that strain Z-245 contains a Vsr-like activity.

(iv) The presence of the GGCC-recognizing *Mth*TI system encoded by pFV1 seems to correlate with the underrepresentation of GGCC-sites in the pFV1 genome. Comparison of the nucleotide sequences of pFV1 and pFZ1 revealed that 15 GGCC-sites present in pFZ1 were changed in pFV1. Remarkably, ten of those changed GGCC-sites in pFV1 displayed the sequence GGTC. The same base substitution would be generated by deamination of a single m^5C nucleotides within the sequence $GG^{mc}CC$ followed by replication, i.e. without previous correction of the G-T mismatch. The observed preference for the replacement of GGCC by GGTC in pFV1 suggests that most of the substitutions have been generated by this mechanism. It therefore seems likely that strain THF has a limited G-T mismatch repair capacity. If so, strain THF should be forced to reduce the number of GGCC-sites available for modification by the m^5C-producing MTase M.*Mth*TI in order to decrease the number of deamination events.

Most advances in physiological, biochemical and genetic analysis of *Methanobacterium* have been made on *M.thermoautotrophicum* strains ΔH and Marburg (for reviews see: 6, 9). The present thesis focused on *M.thermoformicicum* and reported a genetic analysis of plasmids, phages and insertion sequences identified in this species. Since the well-studied methanogen *M.thermoautotrophicum* ΔH turned out to be a non-formate utilizing relative of *M.thermoformicicum*, the here described mobile genetic elements may be instrumental in developing a cloning vector and a transformation system that would allow genetic engineering of strain ΔH.

REFERENCES

1. **Ehrlich, M., M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, and C. W. Gehrke.** (1985). DNA methylation in thermophilic bacteria: N^4-methylcytosine, 5-methylcytosine, and N^6-methyladenine. Nucleic Acids Res. **13**, 1399-1412.
2. **Hennecke, F., H. Kolmar, K. Bründl, and H.-J. Fritz.** (1991). The *vsr* gene product of *E.coli* K-12 is a strand- and sequence specific DNA mismatch endonuclease. Nature (London) **353**, 776-778.
3. **McClelland, M., and A. S. Bhagwat.** (1992). Biased DNA repair. Nature (London) **355**, 595-596.
4. **McClelland, M., and M. Nelson.** (1992). Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res. **20**, 2145-2157.
5. **Phillips, G. J., J. Arnold, and R. Ivarie.** (1987). The effect of codon usage on the oligonucleotide composition of the *E.coli* genome and identification of over- and underrepresented sequences by Markov chain analysis. Nucleic Acids Res. **15**, 2627-2638.
6. **Reeve, J. N.** (1992). Molecular biology of methanogens. Annu. Rev. Microbiol. **46**, 165-191.
7. **Stettler, R., and T. Leisinger** (1992). Physical map of the *Methanobacterium thermoautotrophicum* Marburg chromosome. J. Bacteriol. **174**, 7227-7234.
8. **Touzel, J. P., E. Conway de Macario, J. Nölling, W. M. de Vos, T. Zhilina, and A. M. Lysenko.** (1992). DNA relatedness among some thermophilic members of the species *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol. **42**, 408-411.
9. **Whitman, W. B., T. L. Bowen, and D. R. Boone.** (1992). The methanogenic bacteria. In: The Prokaryotes, 2nd ed., p. 719-767 (eds. A. Balows, H. G. Trüper, M. Dworkin, W. H. Harder, K. H. Schleifer), Springer Verlag, New York.

Samenvatting

De identificatie van de *Archaea* als een derde primaire levensvorm en hun adaptatie aan extreme omstandigheden hebben een aanzienlijke belangstelling gewekt voor de moleculaire biologie van deze organismen. De meeste vooruitgang in het onderzoek van archaeale mobiele genetische elementen, dat wil zeggen virussen, plasmiden en insertie elementen, is gemaakt binnen de groep van halofiele *Archaea*, maar er is weinig bekend over mobiele elementen in methanogenen (zie Hoofdstuk 1 voor een overzicht). Het doel van het in dit proefschrift beschreven onderzoek was om meer inzicht te verkrijgen in de moleculaire genetica van mobiele elementen in methanogenen. Als modelorganismen werd gekozen voor de thermofiele *Methanobacterium* soorten die tot de best bestudeerde methanogenen behoren.

Fylogenetische studies van *M.thermoformicicum* en *M.thermoautotrophicum* staan beschreven in Hoofdstuk 2. Door de vergelijking van twee variable 16S rRNA gebieden was het mogelijk te concluderen dat (i) *M.thermoformicicum* uit twee groepen van stammen bestaat, de zogenaamde Z-245-groep en de CB12-groep die waarschijnlijk tot verschillende species behoren, (ii) *M.thermoautotrophicum* ΔH nauw verwant is met de Z-245-groep, en (iii) *M.thermoautotrophicum* Marburg bij geen van de twee groepen behoort en waarschijnlijk een aparte species vormt. De uit de 16S rRNA analyse bepaalde verwantschap tussen stammen van *M.thermoformicicum* en *M.thermoautotrophicum* kwam overeen met resultaten van DNA-DNA hybridisatie studies (Touzel et al. [1992], *Int. J. Syst. Bacteriol.* 42, 408-411). Beide benaderingen zijn gebaseerd op genotypische eigenschappen die voor de bepaling van fylogenetische verwantschappen in het algemeen betrouwbaarder worden geacht dan fenotypische eigenschappen. De voorgestelde herindeling van de onderzochte thermofiele *Methanobacterium* stammen geeft daarom hun fylogenetische positie waarschijnlijk beter weer dan de indeling die tot nu toe gebruikt is. Een andere, fylogenetisch relevant aspect, gepresenteerd in Hoofdstuk 2, was de ontdekking dat de niet-formiaat gebruikers *M.thermoautotrophicum* ΔH en Marburg sequenties bevatten die overeenkomst vertoonden met de *fdh*A en *fdh*B genen van *M.formicicum*. De aanwezigheid van dergelijke sequenties in de stammen ΔH en Marburg suggereert dat *M.thermoformicicum* en *M.thermoautotrophicum* afkomstig zijn van een gemeenschappelijke, formiaat-gebruikende voorouder.

Hoofdstuk 3 beschrijft de karakterisatie van twee nieuwe archaeale virussen ΦF1 en ΦF3 die in staat zijn verschillende thermofiele *Methanobacterium* stammen te infecteren. Beide fagen verschillen wat betreft hun gastheerspecificiteit en de topologie van hun dubbelstrengs DNA chromosomen. Terwijl ΦF1 een breed gastheerbereik heeft en een lineair, ongeveer 85-kb genoom bezit, is ΦF3 specifiek voor *M.thermoformicicum* FF3 en bezit een circulair, ongeveer 36-kb genoom. Er bestaat geen overeenkomst tussen de genomen van ΦF1 en ΦF3, noch tussen het genomische DNA van verschillende *Methanobacterium* stammen of van ψM1, een faag van *M.thermoautotrophicum* Marburg.

De isolatie en karakterisatie van drie plasmiden, pFV1 (13,5 kb), pFZ1 (11,0 kb) en pFZ2 (11,0 kb), van de *M.thermoformicicum* stammen THF, Z-245 en FTF staat beschreven in Hoofdstuk 4. Door middel van hybridisatie met radioaktief gemarkeerd pFZ1 DNA bleek dat deze plasmiden een familie van nauw verwante elementen vormen. Terwijl pFZ1 en pFZ2 waarschijnlijk identiek zijn, wordt slechts een gedeeltelijke homologie gevonden tussen pFZ1 en pFV1. Daarnaast beschrijft dit hoofdstuk dat hybridisatie met het *fdh*A gen van *M.formicicum* gebruikt kan worden voor 'genetische vingerafdrukken' waardoor het mogelijk is om de species *M.thermoformicicum* in twee groepen te verdelen, de hierboven reeds

genoemde CB12-groep, met de stammen CB12 en SF-4, en de Z-245-groep met de plasmide-bevattende stammen Z-245, FTF en THF en twee nieuwe isolaten FF1 en FF3.

De verwantschap tussen de plasmide DNA's werd bevestigd door de bepaling van de basenvolgorde van pFV1 (13513 bp) en pFZ1 (11014 bp) (Hoofdstuk 5). De vergelijking van deze sequenties liet zien dat de plasmiden pFV1 en pFZ1 in modules zijn georganiseerd: een ruggegraat struktuur, die geconserveerd is op DNA-sequentie niveau, en unieke sequenties met daarin opgenomen plasmide-specifieke elementen.

De organisatie van de *M.thermoformicicum* plasmiden is vergelijkbaar met die van bacteriële chromosomen waarin genetische herrangschikkingen optreden, in het bijzonder inserties en/of deleties, maar waarbij de basisstruktuur wordt bewaard. De grote overeenkomst in DNA volgorde en de vergelijkbare genetische organisatie doen vermoeden dat pFV1 en pFZ1 (en pFZ2) van een gemeenschappelijk voorouder-plasmide afkomstig zijn. In dit geval zouden de plasmiden voor vergelijkbare funkties coderen die van belang zijn voor vermenigvuldiging en handhaving. Deze zouden door de sequenties in de geconserveerde ruggegraatstruktuur gecodeerd kunnen worden. Zulke essentiële plasmidefunkties zouden twee grote palindroom-sequenties kunnen omvatten, alsmede een gen, dat mogelijk voor een NTP-bindend eiwit codeert. Al deze sequenties maken deel uit van de ruggegraatstruktuur die in de plasmiden wordt aangetroffen. In tegenstelling hiermee vertegenwoordigen de plasmide-specifieke sequenties juist additionele elementen die niet-essentiële plasmidefunkties zouden kunnen specificeren.

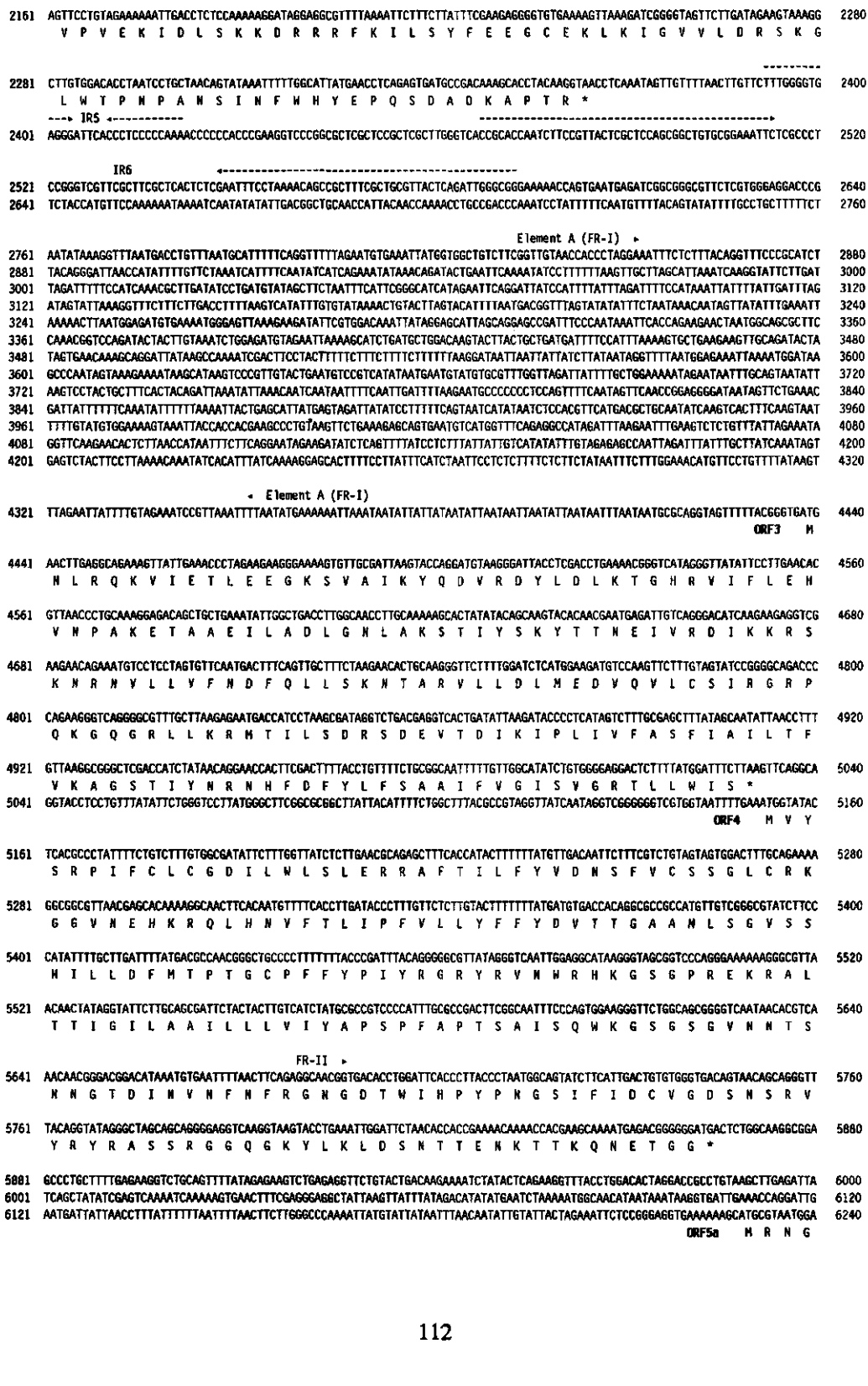
Beide plasmiden pFV1 en pFZ1 bevatten sequenties die overeenkomst vertonen met het chromosomale DNA van verschillende thermofiele methanogenen (Hoofdstuk 6). Een van deze sequenties, FR-I, is een additioneel element van pFV1 en heeft homologe tegenhangers in het chromosomale DNA van verschillende *M.thermoformicicum* stammen en *M.thermoautotrophicum* ΔH. Door vergelijking van de FR-I elementen afkomstig uit pFV1 en chromosomale DNA, kon afgeleid worden dat FR-I een lengte heeft van 1.5 kb en in varianten voorkomt die verschillen in de organisatie van twee subfragmenten. Opvallend was dat overeenkomende subfragmenten van verschillende FR-I elementen volledig identieke nucleotide sequenties bezaten. De FR-I elementen bezitten terminale, direkt herhaalde sequenties en zijn in meerdere kopieën aanwezig. Deze eigenschappen suggereren dat FR-I een nieuw type archaeaal insertie element vertegenwoordigt. Een tweede element, FR-II, maakt deel uit van beide plasmiden pFV1 en pFZ1 en werd aangetroffen in het chromosoom van *M.thermoformicicum* THF, CSM3 en HN4. Uit de DNA sequentie vergelijking van de FR-II elementen, waarvan er twee van plasmide DNA afkomstig waren en een uit het chromosoom, bleek dat de FR-II elementen sterk op elkaar leken en voor een eiwit met nog onbekende eigenschappen zouden kunnen coderen. In tegenstelling tot FR-I werd FR-II maar in een enkele chromosomale kopie gevonden en bezit het geen terminale herhaalde sequenties.

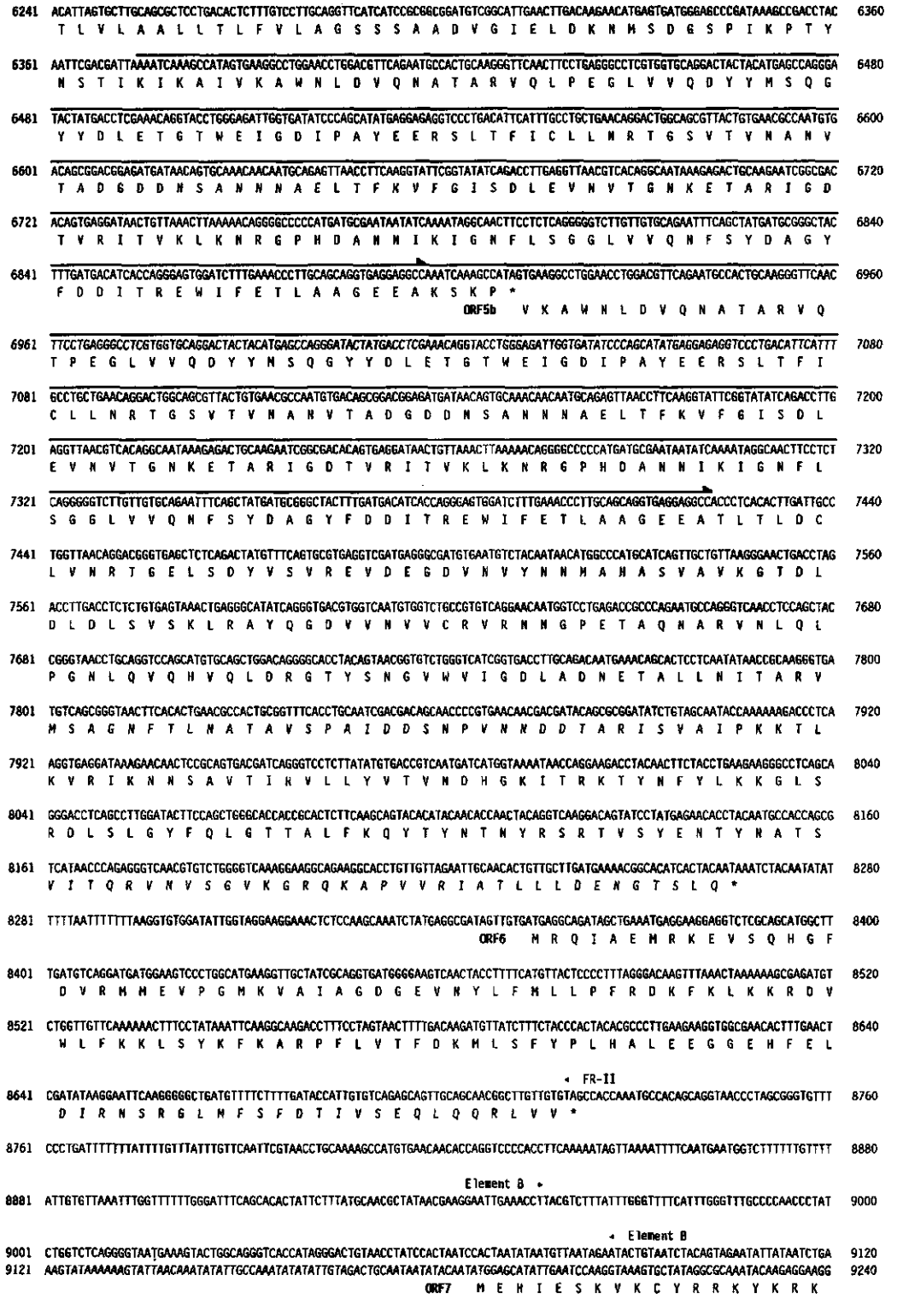
Elk plasmide bevat een additioneel element dat voor componenten van een type II restrictie-modificatie (R/M) systeem codeert: het *Mth*TI systeem van plasmide pFV1 uit stam THF met de herkenningssequentie GGCC en de CTAG-herkennenden R/M systemen *Mth*ZI en *Mth*FI die deel uitmaken van de plasmiden pFZ1 en pFZ2 van de stammen Z-245 en FTF. De gedetailleerde analyse van de systemen *Mth*TI en *Mth*ZI wordt beschreven in respectievelijk Hoofdstuk 7 en Hoofdstuk 8. Opmerkelijk is dat beide R/M modules in een vergelijkbare positie in de ruggegraatstruktuur van beide plasmiden gelokaliseerd zijn. Deze resultaten beschrijven voor het eerst een plasmide-gecodeerde functie in methanogenen.

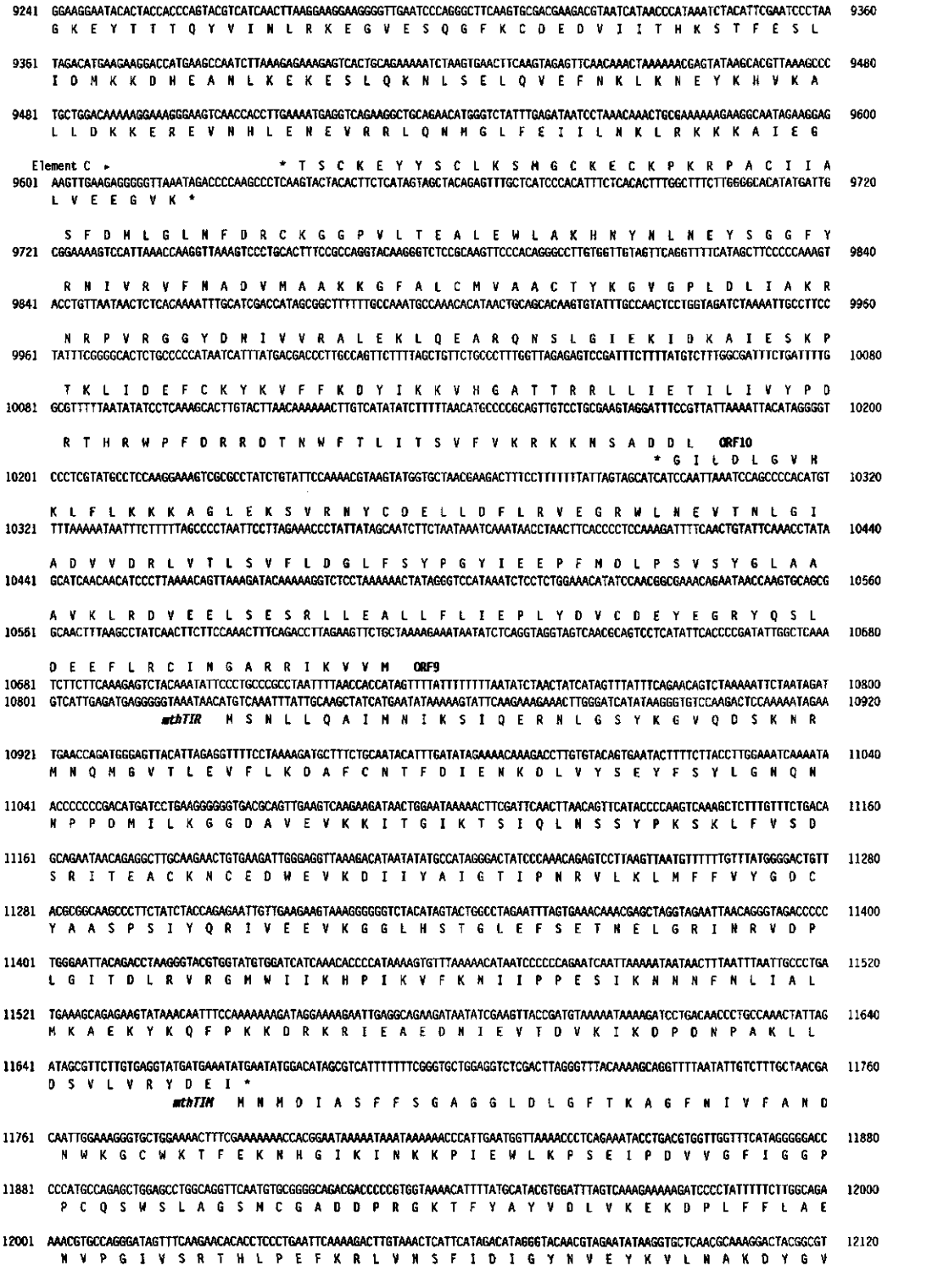
Tenslotte wordt in Hoofdstuk 9 een samenvatting gegeven van de onderzoeksresultaten en worden enkele aanvullende gegevens besproken.

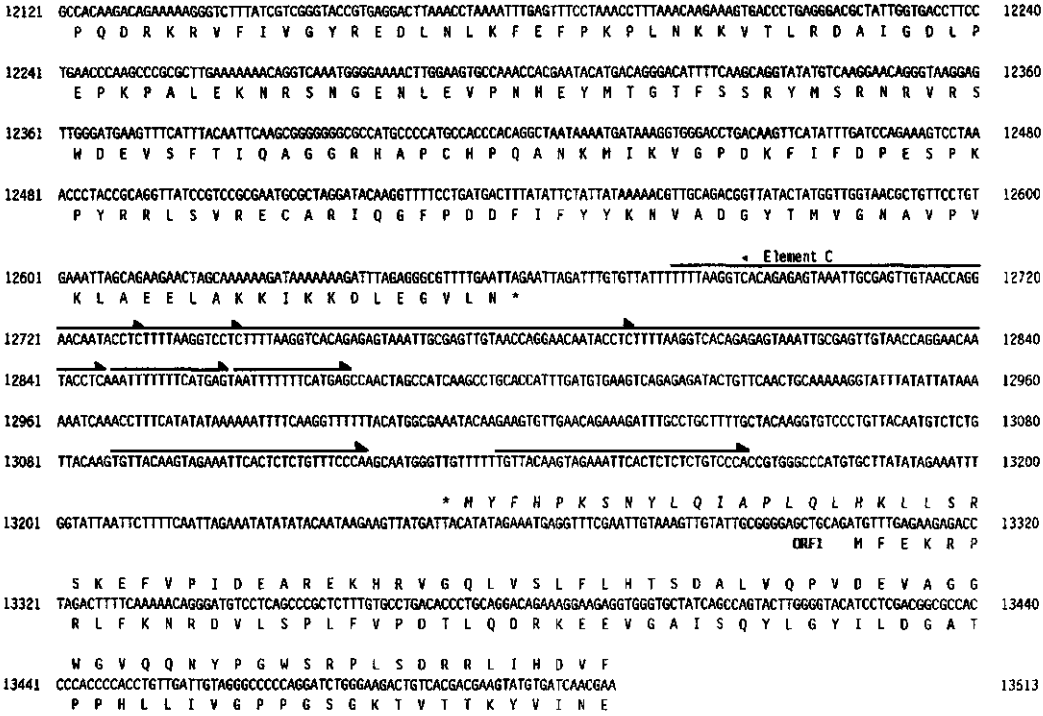
Complete nucleotide sequence of plasmid pFV1 from *M.thermoformicicum* THF (EMBL accession number X68366).
Numbering, open reading frames, insertions, inverted repeats (dashed arrows) and direct repeats (solid arrows) are according to Chapter 5.

```
      KhoI
        E  L  L  V  G  T  V  R  Y  N  L  I  D  N  R  V  A  S  G  V  L  D  C  R  G  N  G  P  G  G  S  A  T  K  P  E  Q  V  N  H
  1 CTCGAGAAGCACACCAGTGACGCGGTAATTGAGTATATCGTTGCGGACGGCACTGCCTACCAGGTCGCAACGTCCATTGCCAGGGCCCCCCGACGCGGTCTTGGGTTCCTGAACATTGTG  120
        L  E  K  H  T  S  D  A  V  I  E  Y  I  V  A  D  G  T  A  Y  Q  V  A  T  S  I  A  R  A  P  R  R  G  L  G  F  L  N  I  V

        L  L  N  P  F  P  G  R  L  A  L  H  Y  N  D  H  V  L  D  I  L  S  K  G  P  V  P  I  F  E  E  V  V  K  R  P  L  R  I  H
121 GAGAAGATTAGGGAAAGGGCCTCGGAGGGCAAGATGATAATTGTCATGGACGAGATCGATAAGACTCTTTCCAGGGACGGGGATAAACTCCTCTACCACCTTTCTAGGGAGCCGAATGTG  240
        E  K  I  R  E  R  A  S  E  G  K  M  I  I  V  M  D  E  I  D  K  T  L  S  R  D  G  D  K  L  L  Y  H  L  S  R  E  P  N  V

        A  N  H  T  K  G  V  L  Q  G  Y  N  I  H  D  A  V  R  P  D  Y  G  R  K  L  W  P  S  D  T  E  G  R  V  A  G  R  L  K  L
241 TGCATTGTGGGTCTTTCCAACAAGTTGACCGTAATGGATATGATCGGCGACTCGGGGGTCATATCCTCGTTTAAGCCAAGGAGAATCAGTTTCGCCCCTTACAGCGCCCCGCAGCTTGAG  360
        C  I  V  G  L  S  N  K  L  T  V  M  D  M  I  G  D  S  G  V  I  S  S  F  K  P  R  R  I  S  F  A  P  Y  S  A  P  Q  L  E

        L  D  Q  V  V  P  N  L  H  R  K  V  V  T  D  Q  F  V  V  N  H  W  Q  T  R  R  E  R  G  L  T  V  P  V  G  P  I  G  K  V
361 GAGATCCTGAACTACAGGGTTGAGATGGCGTTTAACGACGGTGTCCTGGAAGACGACGTTGTGCCACTGTGTGCCGGCGCTCGCGGCCCAGCGTAACGGGGACGCCAGGTATGCCCTTGAC  480
        E  I  L  N  Y  R  V  E  M  A  F  N  D  G  V  L  E  D  D  V  V  P  L  C  A  A  L  A  A  Q  R  N  G  D  A  R  Y  A  L  D

        E  E  A  K  R  R  I  N  R  Y  P  L  K  L  T  N  Y  T  L  R  V  D  P  H  R  S  I  L  N  L  N  L  K  D  P  T  T  D  L  L
481 CTCCTCAGCTTTGCGGCGGATATTGCGATAAGGCAGCTTAAGGGTGTTGTAAGTGAGTCTGACGTCAGGATGGCGACTGATGAGGTTGAGGTTGAGTTTATCAGGCGTAGTATCGAGCAG  600
        L  L  S  F  A  A  D  I  A  I  R  Q  L  K  G  V  V  S  E  S  D  V  R  M  A  T  D  E  V  E  V  E  F  I  R  R  S  I  E  Q

        Q  P  V  I  L  L  D  E  Q  I  G  D  H  G  G  M  P  T  G  G  S  L  Y  I  P  L  V  V  L  H  G  L  L  L  K  P  A  V  G  K
601 TTGAGGGACAATCAGAAGATCCTCCTGTATGCCGTCATGACCTCCCATGGGGGTACCCCCACTGAGATATATAGGAAGTACAACAAGATGACCGAGGAGCAGTTTGGGGGCAACGCCCTT  720
        L  R  D  N  Q  K  I  L  L  Y  A  V  M  T  S  H  G  G  T  P  T  E  I  Y  R  K  Y  N  K  M  T  E  E  Q  F  G  G  N  A  L

        R  L  P  P  K  G  L  E  Q  P  F  K  F  E  V  T  K  N  F  N  L  N  D  T  P  T  T  P  P  P  N  V  P  M   ORF8
721 ACGCAGAGGAGGCTTTCCCAGCTCCTGAGGGAACTTGAACTCTACGGTCTTGTTGAAATTGAGGTTGTCGGTAGGGGTCGTGGGAGGGGGGTTAACTGGCATGTTGTTCCTTCGTCCTCT  840
        T  Q  R  R  L  S  Q  L  L  R  E  L  E  L  Y  G  L  V  E  I  E  V  V  G  R  G  R  G  R  G  V  N  W  H  V  V  P  S  S  S

                                                    (Insert element D in pFZ1)  ▾
841 ATTGACCCTGAACTCATGCTTGAGGCCATAAGGAGGTCCCTCTGAGACTACTTGGGGGCAACTTTCTCTGGGTTTAAGTTGATTGACCTTATACTTGATGGGTACTTTAACTTACCCTTT  960
        I  D  P  E  L  M  L  E  A  I  R  R  S  L  *

961 TACTATTAACTTTTCTTTAATGGGGGTATCTAGGGCTTGTCCTGTAATCTTCCTGAGTGCAGCATTGTCTTTTCAGAACAATTTTTCGTTCAATAAAATTGTTGGGTCTTTGGTTGACCT 1080

                          ----------▸                IR1          ◂---------- -----▸ IR2 ◂-----      ---
1081 CTAGTTTGGTAATGTCCTATGGGTCGGCTAAGCGATTTCCGCCCTTTCGGGTCGCTCGCTTCGCTCGCTCGGAAATCGCCGCCGACCAGGGGGGCTACCACCCCCCACCCCCCGCCCCAC 1200

      --------▸ IR3 ◂----------  ----------------------------------------------------------------------------------------▸
1201 TTTAGGTAACCCATACCTCAACTTGTGATTTTGTGAATTTACCCCTTCACTCCGCCCCGACACTCTTATATAGACGCCAGCTTCGTCAGTACCTCGTATTATACGAAGTAACGCTCGCTT 1320

        IR4 ◂----------------------------------------------------------------------------
1321 CGCTCGCTCTCGCTTCGCTCGTACTTCGTATAATACTAAGAGGTATGACGAAGCTGGCGTGAATTATTTTTGAGTGTCTGGGCTCCGTGAAGGGGTGAACGAACACCACATAAGTGGAGT 1440
1441 TCATTGTAATTAAGTTGTAACTAACCTTTATAAGGATAATATAAGATAGGTGTCTATTAGGGTTGTTAAAATGAATTCAGGTAATCTTTTTTTTTGGGTGGGCAGATGAATCCAGCCAGAA 1560
                                                                ORF2   M  N  S  G  N  L  F  F  G  W  A  D  E  S  S  Q  N

1561 CAAGGGGAGATATCGGGCTATTGGAATGGTAAGCCACCCGGCCGAGTTTGAAAGTAAACTTGAAGGAGATATAAATAACATACTTGACTATGTATTTGATGAATCTGGTTTAAAAAGAAG 1680
        K  G  R  Y  R  A  I  G  N  V  S  H  P  A  E  F  E  S  K  L  E  G  D  I  N  N  I  L  D  Y  V  F  D  E  S  G  L  K  R  R

1681 AGAGTTAAAATGGAATAAAATAGACATATTCCGGTGTAGGGCATATGAGAGGATTGTAGACTACTTTTTAGAGCACACTGGTCTTGGCAATCCACCATTTCGTGTGGATATCTTACGGTG 1800
        E  L  K  W  N  K  I  D  I  F  R  C  R  A  Y  E  R  I  V  D  Y  F  L  E  H  T  G  L  G  N  P  P  F  R  V  D  I  L  R  W

1801 GGACATTGAAGACAGCCGTCATAGTATACAAGGGCGGGATGATAACCAAAACTTGCAGAGGATGTACTATCATCTCTTTTCAAATGTGATTTCTAAGCGTTGGCCTTCTGGAGATTGGTG 1920
        D  I  E  D  S  R  H  S  I  Q  G  R  D  D  N  Q  N  L  Q  R  M  Y  Y  H  L  F  S  N  V  I  S  K  R  W  P  S  G  D  W  C

1921 CTTTTTTCCTGACAAAACGGGTTCAGTCGATTGGGGGGAATTGGCGTCCTTTTTGGATCTTGGCGGTTCAAAAGTAGACCTAAATGAACGGTTTAATATACGAAGTATTAATGAAGTTGA 2040
        F  F  P  D  K  T  G  S  V  D  W  G  E  L  A  S  F  L  D  L  G  G  S  K  V  D  L  N  E  R  F  N  I  R  S  I  N  E  V  D

2041 TTCCAAGGATAATGTACTAGTTCAGGTGGCTGATTTTTTCGCGGGACTCAGTGTTTTTTCTAAGGAGAAGCTTGGGCTTTATGTAGACTGGAAGTTTGAAAAACGTGGACAGCAAAGACT 2160
        S  K  D  N  V  L  V  Q  V  A  D  F  F  A  G  L  S  V  F  S  K  E  K  L  G  L  Y  V  D  W  K  F  E  K  R  G  Q  Q  R  L
```

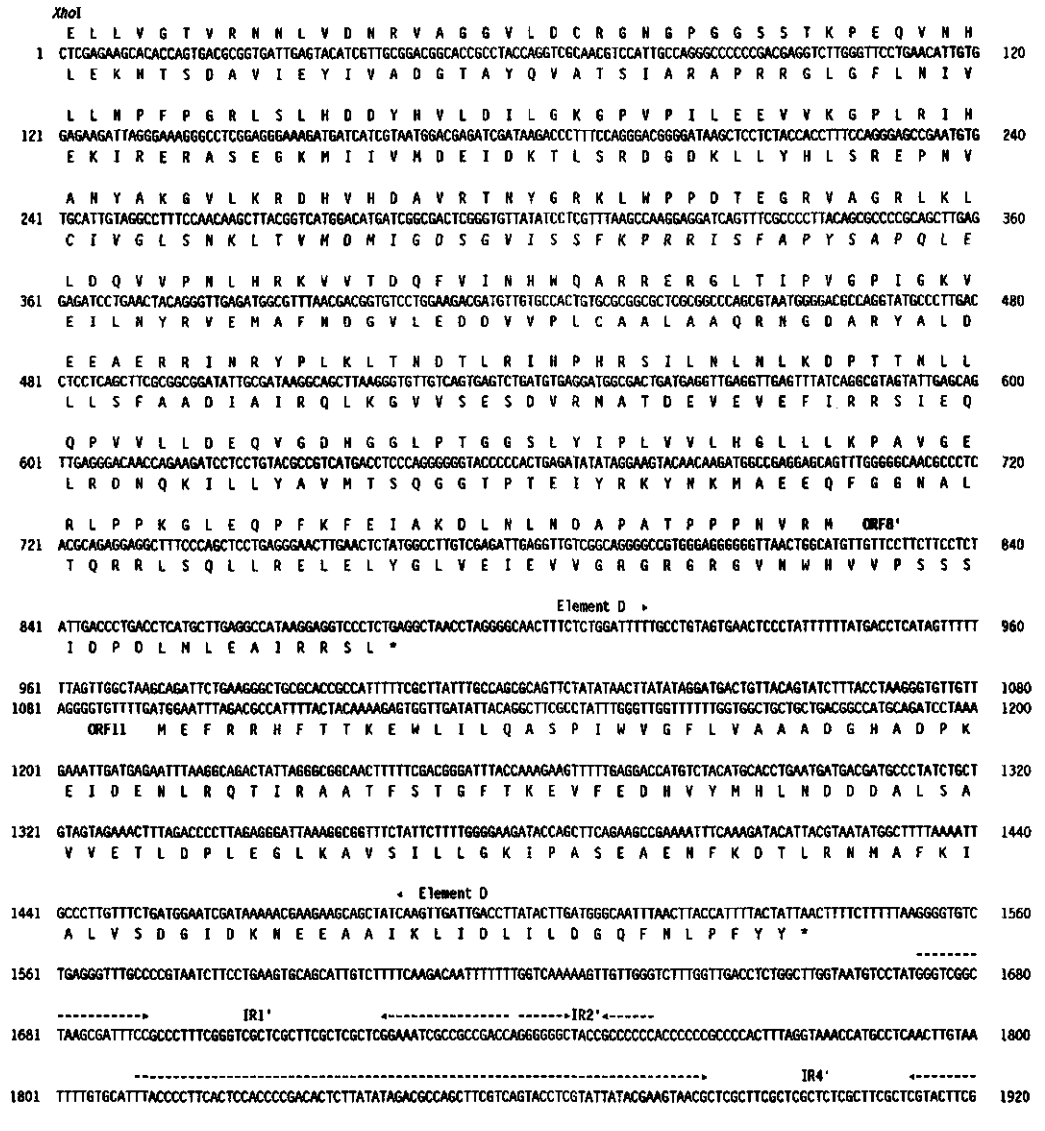

IR5
IR6
Element A (FR-I)
ORF3
ORF4
FR-II
ORF5a


ORF5b
FR-II
Element B
ORF7

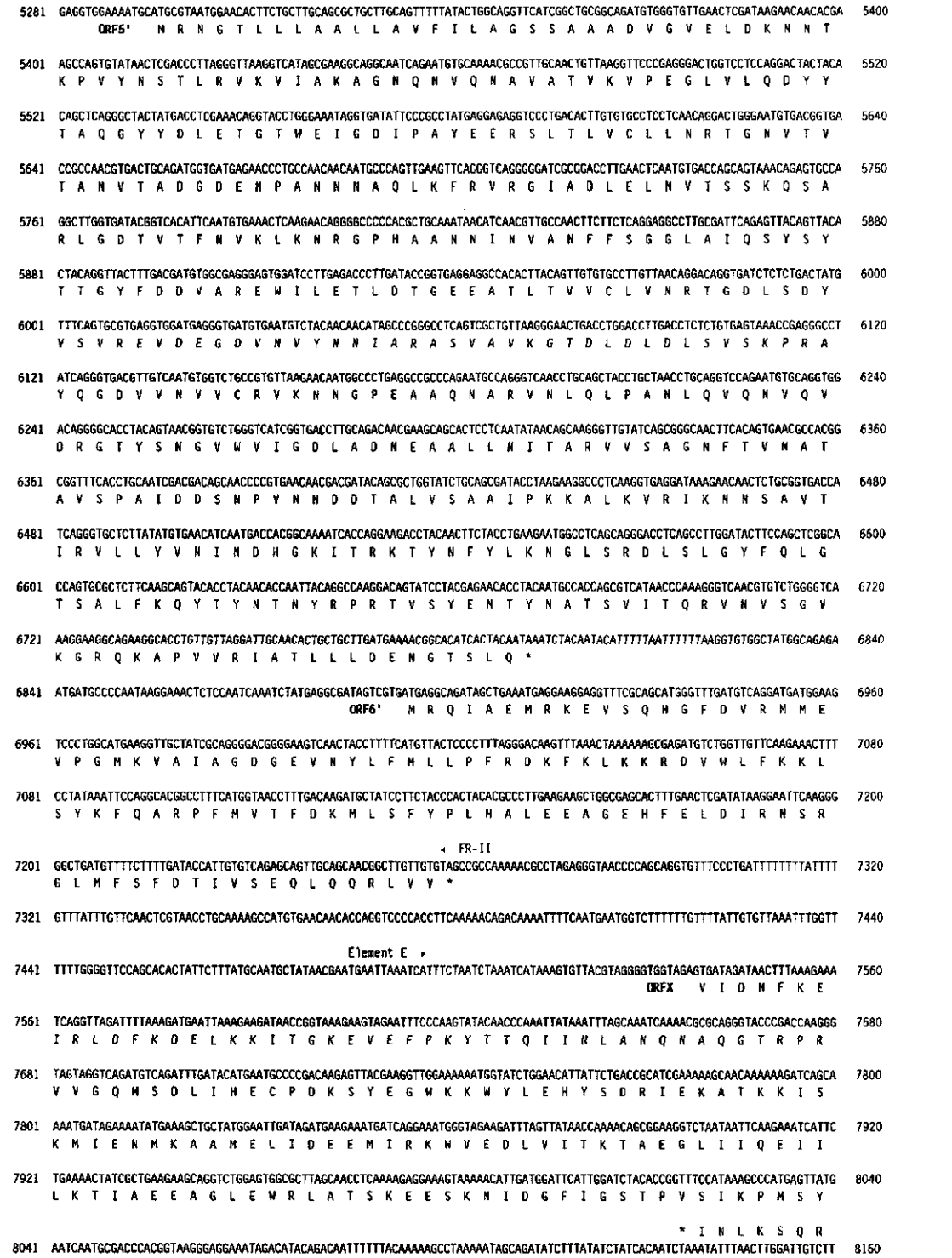

Element C
ORF10
ORF9
mthTIR
mthTIM

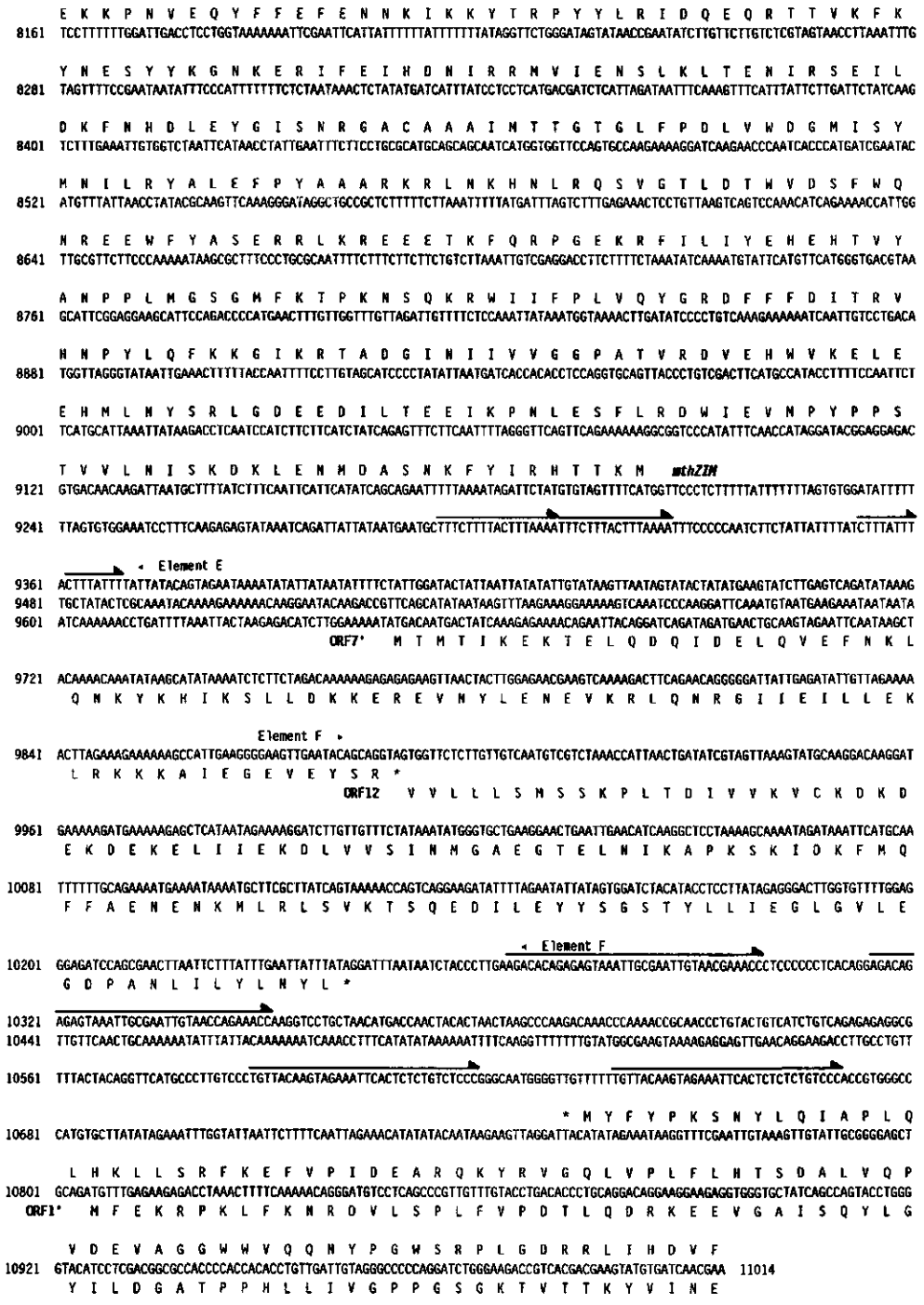

Element C
ORF1

Complete nucleotide sequence of plasmid pFZ1 from *M.thermoformicicum* Z-245 (EMBL accession number X68367).
Numbering, open reading frames, insertions, inverted repeats (dashed arrows) and direct repeats (solid arrows) are according to Chapter 5.



2161 TATCGGGCTATCGGAATGGTAAGCCATCCGGCTGAGTTTGGAAGTGAACTTGAAGGAGATATAAATAGTATACTTGACTCGGTCGTTGATGAATCTGGTTTAAAAAGAAGAGAGTTGAAA 2280
Y R A I G M V S H P A E F G S E L E G D I N S I L D S V V D E S G L K R R E L K
2281 TGGAATAAAATAGACATATTCCGGTGTAGAGCATATGAGAAGATTGTAGACTACTTTTTAGAACTCAGTAACCTTGGTAATCCACCAGTTCGTGTGGATATCTTACGGTGGGACATTGAA 2400
W N K I D I F R C R A Y E K I V D Y F L E L S N L G N P P V R V D I L R W D I E
2401 GACAGCCGTCACAGTATACAAGGGCGGGATGATAACCAAAACCTGCAGAGGATGTACTATCATCTATTTTCCAATGTGATTTCTAAGCGTTGGCCTTCTGGAGATTGGTGCTTTTTCCCT 2520
D S R H S I Q G R D D N Q N L Q R M Y Y H L F S N V I S K R W P S G D W C F F P
2521 GATGAAACAGGTTCAGTAGATTGGGAGGAAGTGGCGTTCTTCTTGGATATTGGTGGTTCAAAGGTAGACCTAAAGGAACAGTTTAATATACTAAGTATTAATGAAGTTGATTCCAAGGAT 2640
D E T G S V D W E E V A F F L D I G G S K V D L K E Q F N I L S I N E V D S K D
2641 AATGTACTAGTTCAAGTGGCTGACTTTTTCGCGGGACTCAGTGTTTTTTCTAAGGAGAAGTTTGGGCTTTATGTAGACTGGAACTACGAAGAATGTGGACAGCAAAGATTAGTTCCTGTA 2760
N V L V Q V A D F F A G L S V F S K E K F G L Y V D W N Y E E C G Q Q R L V P V
2761 GAAAACGTTGAGTTCTCCAAAAAGGATAGGAAGCGTTTTGAAATTCTTTCTTATTTCAACAAGAGGTGCAAAGAGTTAAAGATGGGGGTAAGTCTTGATACAAGTGAAGGCTTGTGGACA 2880
E N V E F S K K D R K R F E I L S Y F N K R C K E L K M G V S L D T S E G L W T
IR5'
2881 CCTAAACCTGCTAACAGTATAAATTTCTGGCATTATGAACCCCAGAGTGATGCCGACAAAGCACCTATAAGGTAACCTCAAATAGTTGTTTTTAACTTGTTCTTTGGGGTGAGGGATTCA 3000
P K P A N S I N F W H Y E P Q S D A D K A P I R *
IR6'
3001 CCCTCCCCCCAAACCCCCCACCCGAAGGCCCCGGCGCTCGCTACGCTCGCTTGGGTCACCGCACCAATCTTCCGTTACTCGCTCCAGCGGCTGTGCGGAAATTCTCGCCCTTCGGGCCGT 3120
3121 TCGCTTCGCTCACTCTCGAATTTCCTAAAACAGCCGCTTTCGCTGCGTTACTCAGATTGGGCGGGAAAAACCAGTGAATGAGATTGGCGGGCGTTCTCGTGGGGGCCCCGTCCACCCTGT 3240
3241 TCCAAAAAATAAAATCAATATATATTGACGGCTGCAACCATTACAACCAAAACCTGCCGACCCAAATCCTATTTTTTCAATGTTTTACAGTATATTTTGCCTTCTTTTTCTAATATAAAG 3360
(Insertion element A [FR-I] in pFV1) ******
3361 GTTTAATGACATGTTTAATGCATTTTTCAGGTTTTTAGAATGTTAAATTATGGCGGCTGTCTTCGGTTGTAACCGCACCCCATGAGATTTTAATATGAAAAAATTAAATAATATTATTAT 3480
ORF3' M K K L N N I I M
3481 GATATTAATAATGCACAGGTACTTCTTACGGGTGATGAACTTGAGGCAGAAAGTTATTGAGGCCCTAGAAGAAGGGAAGAGTGTTGCGATTAAGTACCAGGATGTAAGGGATTACCTCGA 3600
I L I N H R Y F L R V N N L R Q K V I E A L E E G K S V A I K Y Q D V R D Y L D
3601 CCTGAAAACAGGTCATAGGGTTATATTCCTTGAACACATTAACCCTGCAAAGGAGACCGCCGCTGAAATATTGGCTGACCTTGGCAACCTTGCAAAAAGCACTATATACAGCAAGTATAC 3720
L K T G H R V I F L E H I N P A K E T A A E I L A D L G N L A K S T I Y S K Y T
3721 AACGAATGAGATTGTCAGGGACATCAAGAAGAGGTCGAAGAACAGGAATGTCCTCCTGGTGTTCAATGACTTTCAGTTGCTTTCTAAGAACACTGCAAGGGTTCTTTTGGATCTCATGGA 3840
T N E I V R D I K K R S K N R N V L L V F N D F Q L L S K N T A R V L L D L M E
3841 AGATGTCCAAGTTCTTTGTAGTATCAGGGGCAGACCCCAGAAGGGTCAGGGGCGTTTGCTTAAGAGAATGACCATCCTAAGCGATAGGTCTGACGAGGTCACTGATATTAAGATACCCCT 3960
D V Q V L C S I R G R P Q K G Q G R L L K R M T I L S D R S D E V T D I K I P L
3961 CATAGTCTTTGCGAGCTTTATAGCAATATTAACCTTTGTTAAGGCGGGCTCGGCTATCTATAACAGGAACCACTTCGACTTTTACCTGTTTTCTGCGGCAATTTTTGTTGGCATATCTGT 4080
I V F A S F I A I L T F V K A G S A I Y N R N H F D F Y L F S A A I F V G I S V
4081 GGGGAGGACTCTTTTATGGATTTCTTAAGTTCAGGCAGGTGCCTCCTGTTTATCTTCTGGGTCGGCCTTCGGCGCGGCTTATTACATTTTCTGGCTTTACGCCGCAGGTTATCAGTAAGT 4200
G R T L L W I S *
4201 CGGTGGGGGTCGTGGTAATTTGAAATGGTATACGCACGCCCTATTTTCGGTCTTTGTGGCGGTATTCTTTGGTTATCTCCTGAACGCAGATCTTTCACCATACTTTTTTATGTTGACAAT 4320
ORF4' M V Y A R P I F G L C G G I L W L S P E R R S F T I L F Y V D N
4321 TCTTTCATCTGTAGTGGTCGACTTTGCAGAAAAGGCGGCGTTAACGAGCACAAAAGGCAACTTCACAATGTTTTCACCCTGATACCCTTTGTTCTCTTGTACCTCTTTTATGATGTGACC 4440
S F I C S G R L C R K G G V N E H K R Q L H N V F T L I P F V L L Y L F Y D V T
4441 ACAGGCGCCGCCATGTTGTCGGGCGTATCTTCCCATATTTTGCTTGATTTCATGACTCCAACGGGCTGCCCCTTTTTTTACCCGATTTACAAGGGGCGTTACAGGGTCGATTGGAGGCAT 4560
T G A A M L S G V S S H I L L D F M T P T G C P F F Y P I Y K G R Y R V D W R H
4561 AAGGGTAGTGGTCCCAGGGAAAAAGGGCGTTAACAACTATAGGTATTCTTGCAGCGATTCTACTCCTTGTTATCTATGCGCCGTCCCCATTTGCGCCGACTTCGGCGATTTCCCAGTGG 4680
K G S G P R E K R A L T T I G I L A A I L L L V I Y A P S P F A P T S A I S Q W
FR-II ▸
4681 AAGGGTTCTGGCAGCGGGGCCTCAAACAACAGGACGGACATAAATGTGAATTTTAACTTCAGAGGCAACGGTGACACCTGGATTCACCCTTACCCTAATGGCAGTATCTTCATTGACTAT 4800
K G S G S G A S N N R T D I N V N F N F R G N G D T W I H P Y P N G S I F I D Y
4801 GTGAGCGACGGTGACAGCAGGGTTTACAGGTACAGGGGTATCAGCCGGGGAGGGCAGGGAAAGCACCTGAAATTGGATTCTAACACCACCGAAAACAAAACCACGAAGCAAAATGAGACG 4920
V S D G D S R V Y R Y R G I S R G G Q G K H L K L D S N T T E N K T T K Q N E T
4921 GGGGGATGACTCTGGCAAGGCGGAGCCCTGCTTTTGAGAAGGTATGCAGTTTTATAGAGAAGTCTGAGAGGTTCTGTACTGACAAGAAAATCTATACTCAGAAGGTTTACCTGGACACCA 5040
G G *
5041 GGACCGCCTGTAAGCTTGAGATTATCAGCTATATCGAGTCAAAGTCAAAAAGTGAACTTTCGAGGGAGGCTATTAAGTTATTTATAGACATATATGAATCTAAAAATGGCAACATAATAA 5160
5161 ATAAGGTGATTGAAAGCAGGATTGAATGATTATTAACCTTTATTTTTTAATTTTAACTTCTTTTTGGGCCTAAAATTATGTATTATAATTTAACAATATTATATTACTAGAATTCTCCAG 5280


ORF5'
FR-II
Element E
ORF6'
ORFX


mthZIN
Element E
ORF7'
Element F
ORF12
ORF1'

Complete nucleotide sequence of cFR-II from *M.thermoformicicum* CSM3 (EMBL accession number X69114; see also Chapter 6).

```
    EcoRI                                   cFR-II ▸
  1 GAATTCACAAACCCAAGCCTCCATCTCAGGATCACCAACCCACAGAGGGACACCTGGGTCCACCCTTTCCCCAACGGCAGCTACTTCATTGACGTGAGCGACTCCCCAGGGACTGATAGA  120
121 ACAGTTTATCACAGGTACTGCACTAAGCCGGGGCCGGAGCGTAAGGTTAGGTCCTCTGAGGGCCAGAGCAATGGGAGTGCTGGAGAAAACGGTACGGGGTGATGTTCTGTCGGGATCTGA  240
241 TATCGATGATGTGGTGAAGTTCATATCGAAGCTGGATTATGATCACGGCAAGAAGATATTCCTCCAGAGGGTTTACCTTGACCTCAACACTGCCTGTAAACTGCATATACTCAGCTTCAT  360
361 TGAGGAGAGGTCCAGGAGTGACATCTCGCGGAGGGCTATTCAGATGATGATAGATGATTATGAGAGGGAATACGGGAACCTCATTAACCTTGTGAAGAGGGACATCCTCATGTGGATATA  480
481 GGGCCCTTGCAGCGGTTCCATCCCCATTTATTTTTTCTTATGTAACTTGTCCACTGATCAGCATCTCCATGGTTACATAAGAATGTACCCAAAGGCATGCTTTAAAGCCCCACCCGAGTT  600
601 TCCAGTGTGAATAACATAGGAGGTGAAACATGCGAAGTGCTAAATTCGTTCTTGCAGCTCTCCTGGCCATCTTTGTTCTTGCAGGTTGATCGGCTGCAGCGGATGTAGGCATTGAACTTG  720
721 ACAAGAACATGACCGACGGGAGCGCAATAAAGCCGACGTACAATTCGACGATGAAAATCAAGGCCATAGTGAAGGCTGGTAACCAGGACGTTCAGAATGTGACTGCAACGGTACGACTGC  840
    cORF5  M T D G S A I K P T Y N S T M K I K A I V K A G N Q D V Q N V T A T V R L

841 CTGAGGGCCTCGTACTCCAGGACTACTACATGAGCCAGGGATACTATGACCTCGAGACAGGTACCTGGGAGGTTGGTGACATCCCTGCCTATGAGGAGAGGTCCCTGACATTCGTTTGCC  960
    P E G L V L Q D Y Y M S Q G Y Y D L E T G T W E V G D I P A Y E E R S L T F V C

961 TGCTGAACAGGACTGGCAGCATTACTGTGAACGCCAGTGTAACAGCTGATGGGGATGATAACAGTGCAAACAACAATGCAGAGTTAGGCTTCAAGGTGTTTGGTATATCAGACCTTCAGC 1080
    L L N R T G S I T V N A S V T A D G D D N S A N N N A E L G F K V F G I S D L Q

1081 TTAACGTCACGAGCAATAAACAGACCCCACGGCTCGGCGAAACTGTGAGACTGACGGTTAAACTTAAAAACAATGGCCCCCATGACGCGAACAATGTCAAGGTAGGTAACTTCCTCTCAG 1200
    L N V T S N K Q T P R L G E T V R L T V K L K N N G P H D A N N V K V G N F L S

1201 GCGGACTCGTCGTGCAGAGTTTCAGCTATGATGCCGGCTACTTCGATGACCTGACAAAGGAGTGGGTCTTTGACACCCTTGCAGCGAGTGAAGAGGCGACCCTCACACTTGATTGCCTTG 1320
    G G L V V Q S F S Y D A G Y F D D L T K E W V F D T L A A S E E A T L T L D C L

1321 TCAACAGGACAGGTGAACTCTCAGACTATGTTTCAGTAAGGGAGGTCGATGAGGGCGACACCAACGTTTATAACAACATGGCCCGCGCCACCCTCTCGGTGAGGGGCACTGACCTCGACC 1440
    V N R T G E L S D Y V S V R E V D E G D T N V Y N N M A R A T L S V R G T D L D

1441 TCGACGTTTCAGCGAGTAAGACAAGGGCCTATCAGGGTGATGTGATCAATGTTATCTGCCGCGTGAAGAACAATGGCCCTGAGAATGCCCAGAACGTGATTGTCAACCTGCAGCTACCCG 1560
    L D V S A S K T R A Y Q G D V I N V I C R V K N N G P E N A Q N V I V N L Q L P

1561 TGAACCTACAGGTCCAGAACGTCCAGGTGGACAGGGGTACTTACAGTAACGGCAACTGGTCCATCGGTGATCTTGCAGACAATGAAACAGCGCTTCTCAACATCACTGCAAGGATTGTGT 1680
    V N L Q V Q N V Q V D R G T Y S N G N W S I G D L A D N E T A L L N I T A R I V

1681 CTGCAGGTAACTTCACAGTGAACGCCTCTGCAGTCTCATCTGCAGTTGATGACAGTAACTCCGTGAACAACGATGATACAGCGATGATATCTGCAGCAGTCCCCAAGAGGGCCCTCAAAC 1800
    S A G N F T V N A S A V S S A V D D S N S V N N D D T A M I S A A V P K R A L K

1801 TCAAGATAAAGAACAACTCTGCAGTGACGATCAGAGTACTCCTATATGTTACAGTCAATGATCACGACAAAATCACCAGGAAGACCTACAACTTCTACCTCAAGAAGGGCCTGAGCAGAG 1920
    L K I K N N S A V T I R V L L Y V T V N D H D K I T R K T Y N F Y L K K G L S R

1921 AACTCAACCTTGGGTACTTCCAGATCGGTACAAGCGCCCTCTTCAAGCAGTACACCTACAACACCAACTACAGGTCAAGGACAGTGGCCTATGAGAACATCTACAACTCCACCGGTGTCA 2040
    E L N L G Y F Q I G T S A L F K Q Y T Y N T N Y R S R T V A Y E N I Y N S T G V

2041 GGACACAGAGGGTGAGCGTCAGTGGTGTGAAAGGAAGGCAGAAGGCACCTGTGGTGAGGCTAACAACCTTAACGTTCGACGAAAACGGTTCAAACATACAATAAACAACTTTCAATTTAT 2160
    R T Q R V S V S G V K G R Q K A P V V R L T T L T F D E N G S N I Q *

2161 TTTTTTGGGAGGTAAATTATGGATATAGAAGATGCTAAAAAGATGGCTTCCCTGCTGGAAATTAAAAATGAGCTAAAGGAATATGATTTTGAAATGAGGATCTTAAAGGATGAAGTGGAT 2280
2281 ATTGCAATAGCTGGAGATGGTGAAACGACATATCTATTCATAATGTTATTGCCTTACAGGAAGAAGTTCAGTATACGTAAAAGACATGTCTGGAAGTTCAAGACCCTTGCCTATAAGTTT 2400
2401 AAGGCAAACCATACATTTTGACCTACAATATCATGACAACGATATATCCTTTACATGCACTTAATGACGCAGGTGATCATTTCGAATGGATGTTGAAAAGGCGAAGGGATTAATGTTCTC 2520

                                        ◂ cFR-II
2521 TTTTGGTACAATAGTCTCAGAAGAATTGCAGGAAAGACTTGCAGTTTGAACATTTATTTTTTTGAGGTCCTCAGAGTTGACTACGGTGCTGTTGACAAAGTCAAGGCAGTTATGGACAGT 2640
2641 TTAGTGATCAAAGGTTTTGATCATCACAAATCAAA                                                                                    2675
```

Curriculum vitae

The author of the present thesis was borne in December, 1960, in Clausthal-Zellerfeld, Germany. After having completed the Albert-Schweitzer Gymnasium in Alsfeld, Germany, he started in 1980 the study of biology at the Philipps-University of Marburg, Germany. The graduation in the disciplines of molecular genetics (Prof. Klein), microbiology (Prof. Thauer), biochemistry (Prof. Kindl) and botany (Prof. Henssen) in 1985 was followed by a 'Diplomarbeit' at the lab of Prof. Klein and Prof. Thauer, which was finished in November 1986. In January 1988 he started as a Ph.D. student at the Department of Microbiology, Bacterial Genetics Group, Wageningen Agricultural University, The Netherlands.

Publication list

- **Nölling, J., M. Frijlink, and W. M. de Vos.** (1991). Isolation and characterization of plasmids from different strains of *Methanobacterium thermoformicicum*. J. Gen. Microbiol. **137**, 1981-1986.

- **Nölling, J., and W. M. de Vos.** (1992). Characterization of the archaeal, plasmid-encoded restriction-modification system *Mth*TI from *Methanobacterium thermoformicicum* THF: homology to the bacterial *Ngo*PII system from *Neisseria gonorrhoeae*. J. Bacteriol. **174**, 5719-5726.

- **Nölling, J., and W. M. de Vos.** (1992). Identification of the CTAG-recognizing restriction-modification systems *Mth*ZI and *Mth*FI in *Methanobacterium thermoformicicum* and characterization of the plasmid-encoded *mthZIM* gene. Nucleic Acids Res. **20**, 5047-5052.

- **Nölling, J., F. J. M. van Eeden, R. I. L. Eggen, and W. M. de Vos.** (1992). Modular organization of related archaeal plasmids encoding different restriction-modification systems in *Methanobacterium thermoformicicum*. Nucleic Acids Res **20**, 6501-6507.

- **Touzel, J. P., E. Conway de Macario, J. Nölling, W. M. de Vos, T. Zhilina, and A. M. Lysenko.** (1992). DNA relatedness among some thermophilic members of the genus *Methanobacterium*: emendation of the species *Methanobacterium thermoautotrophicum* and rejection of *Methanobacterium thermoformicicum* as a synonym of *Methanobacterium thermoautotrophicum*. Int. J. Syst. Bacteriol. **42**, 408-411.

- **Nölling, J., D. Hahn, W. Ludwig, and W. M. de Vos.** (1993). Phylogenetic analysis of thermophilic *Methanobacterium* sp.: evidence for a formate-utilizing ancestor. System. Appl. Microbiol., in press.

- **Nölling, J., F. J. M. van Eeden, and W. M. de Vos.** (1993). Distribution and characterization of plasmid-related sequences in the chromosomal DNA of different thermophilic *Methanobacterium* strains. Mol. Gen. Genet., accepted for publication.

- **Nölling. J., A. Groffen, and W. M. de Vos.** (1993). ΦF1 and ΦF3: two novel virulent, archaeal phages infecting different thermophilic strains of the genus *Methanobacterium*. Submitted for publication.